Extremum Problems with Total Variation Distance and their Applications

Charalambos D. Charalambous, Ioannis Tzortzis, Sergey Loyka and Themistoklis Charalambous

Abstract

The aim of this paper is to investigate extremum problems with pay-off being the total variational distance metric defined on the space of probability measures, subject to linear functional constraints on the space of probability measures, and vice-versa; that is, with the roles of total variational metric and linear functional interchanged. Utilizing concepts from signed measures, the extremum probability measures of such problems are obtained in closed form, by identifying the partition of the support set and the mass of these extremum measures on the partition. The results are derived for abstract spaces; specifically, complete separable metric spaces known as Polish spaces, while the high level ideas are also discussed for denumerable spaces endowed with the discrete topology. These extremum problems often arise in many areas, such as, approximating a family of probability distributions by a given probability distribution, maximizing or minimizing entropy subject to total variational distance metric constraints, quantifying uncertainty of probability distributions by total variational distance metric, stochastic minimax control, and in many problems of information, decision theory, and minimax theory.
Keywords: Total variational distance, extremum probability measures, signed measures.

I. Introduction

Total variational distance metric on the space of probability measures is a fundamental quantity in statistics and probability, which over the years appeared in many diverse applications. In

C. D. Charalambous and I. Tzortzis are with the Department of Electrical Engineering, University of Cyprus, Nicosia, Cyprus. Emails: {chadcha,tzortzis.ioannis}@ucy.ac.cy.

S. Loyka is with the School of Information Technology and Engineering, University of Ottawa, Ontario, Canada. Email: sergey.loyka@ieee.org.

T. Charalambous is with the School of Electrical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden. Email: themisc@kth.se.

information theory it is used to define strong typicality and asymptotic equipartition of sequences generated by sampling from a given distribution [1]. In decision problems, it arises naturally when discriminating the results of observation of two statistical hypotheses [1]. In studying the ergodicity of Markov Chains, it is used to define the Dobrushin coefficient and establish the contraction property of transition probability distributions [2]. Moreover, distance in total variation of probability measures is related via upper and lower bounds to an anthology of distances and distance metrics [3]. The measure of distance in total variation of probability measures is a strong form of closeness of probability measures, and, convergence with respect to total variation of probability measures implies their convergence with respect to other distances and distance metrics.

In this paper, we formulate and solve several extremum problems involving the total variational distance metric and we discuss their applications. The main problems investigated are the following.

(a) Extremum problems of linear functionals on the space of measures subject to a total variational distance metric constraint defined on the space of measures.

(b) Extremum problems of total variational distance metric on the space of measures subject to linear functionals on the space of measures.

(c) Applications of these extremum problems, and their relations to other problems.

The formulation of these extremum problems, their discussion in terms of applications, and the contributions of this paper are developed at the abstract level, in which systems are represented by probability distributions on abstract spaces (complete separable metric space, known as Polish spaces [4]), pay-offs are represented by linear functionals on the space of probability measures or by distance in variation of probability measures, and constraints by linear functionals or distance in variation of probability measures. We consider Polish spaces since they are general enough to handle various models of practical interest.

Utilizing concepts from signed measures, closed form expressions of the probability measures are derived which achieve the extremum of these problems. The construction of the extremum measures involves the identification of the partition of their support set, and their mass defined on these partitions. Throughout the derivations we make extensive use of lower and upper bounds of pay-offs which are achievable. Several simulations are carried out to illustrate the different features of the extremum solution of the various problems. An interesting observation concerning

one of the extremum problems is its equivalent formulation as an extremum problem involving the oscillator semi-norm of the pay-off functional. The formulation and results obtained for these problems at the abstract level are discussed throughout the paper in the context of various applications, often assuming denumerable spaces endowed with the discrete topology. Some specific envisioned applications of the theory developed are listed below.

(i) Dynamic Programming Under Uncertainty, to deal with uncertainty of transition probability distributions, via minimax theory, with total variational distance metric uncertainty constraints to codify the impact of incorrect distribution models on performance of the optimal strategies [5]. This formulation is applicable to Markov decision problems subject to uncertainty.

(ii) Approximation of Probability Distributions with Total Variational Distance Metric, to approximate a given probability distribution μ on a measurable space $(\Sigma, \mathscr{B}(\Sigma))$ by another distribution ν on $(\Sigma, \mathscr{B}(\Sigma))$, via minimization of the total variational distance metric between them subject to linear functional constraints. Model and graph reduction can be handled via such approximations.

(iii) Maximization or Minimization of Entropy Subject to Total Variational Distance Metric Constraints, to invoke insufficient reasoning based on maximizing the entropy $H(\nu)$ of an unknown probability distribution ν on denumerable space Σ subject to a constraint on the total variational distance metric.

The rest of the paper is organized as follows. In section II, total variational distance is defined, the extremum problems are introduced, while several related problems are discussed together with their applications. In section III, some of the properties of the problems are discussed. In section III-A, signed measures are utilized to convert the extremum problems into equivalent ones, and to characterize the extremum measures on abstract spaces. In section IV, closed form expressions of the extremum measures are derived for finite alphabet spaces. In section V, the relation between total variational distance and other distance metrics is discussed. Finally, in section VI several examples are worked out to illustrate how the optimal solution of extremum problems behaves by examining different scenarios concerning the partition of the space Σ.

II. Extremum Problems

In this section, we will introduce the extremum problems we shall investigate. Let (Σ, d_Σ) denote a complete, separable metric space and $(\Sigma, \mathscr{B}(\Sigma))$ the corresponding measurable space, where $\mathscr{B}(\Sigma)$ is the σ-algebra generated by open sets in Σ. Let $\mathscr{M}_1(\Sigma)$ denote the set of probability measures on $\mathscr{B}(\Sigma)$. The total variational distance[1] is a metric [6] $d_{TV} : \mathscr{M}_1(\Sigma) \times \mathscr{M}_1(\Sigma) \to [0, \infty)$ defined by

$$d_{TV}(\alpha, \beta) \equiv ||\alpha - \beta||_{TV} \stackrel{\triangle}{=} \sup_{P \in \mathscr{P}(\Sigma)} \sum_{F_i \in P} |\alpha(F_i) - \beta(F_i)| \,, \tag{1}$$

where $\alpha, \beta \in \mathscr{M}_1(\Sigma)$ and $\mathscr{P}(\Sigma)$ denotes the collection of all finite partitions of Σ. With respect to this metric, $(\mathscr{M}_1(\Sigma), d_{TV})$ is a complete metric space. Since the elements of $\mathscr{M}_1(\Sigma)$ are probability measures, then $d_{TV}(\alpha, \beta) \leq 2$. In minimax problems one can introduce an uncertainty set based on distance in variation as follows. Suppose the probability measure $\nu \in \mathscr{M}_1(\Sigma)$ is unknown, while modeling techniques give access to a nominal probability measure $\mu \in \mathscr{M}_1(\Sigma)$. Having constructed the nominal probability measure, one may construct from empirical data, the distance of the two measures with respect to the total variational distance $||\nu - \mu||_{TV}$. This will provide an estimate of the radius R, such that $||\nu - \mu||_{TV} \leq R$, and hence characterize the set of all possible true measures $\nu \in \mathscr{M}_1(\Sigma)$, centered at the nominal distribution $\mu \in \mathscr{M}_1(\Sigma)$, and lying within the ball of radius R, with respect to the total variational distance $||\cdot||_{TV}$. Such a procedure is used in information theory to define strong typicality of sequences. Unlike other distances used in the past such as relative entropy [7]–[11], quantifying uncertainty via the metric $||\cdot||_{TV}$ does not require absolute continuity of measures[2], i.e., singular measures are admissible, and hence ν and μ need not be defined on the same space. Thus, the support set of μ may be $\tilde{\Sigma} \subset \Sigma$, hence $\mu(\Sigma \setminus \tilde{\Sigma}) = 0$ but $\nu(\Sigma \setminus \tilde{\Sigma}) \neq 0$ is allowed. For measures induced by stochastic differential equations (SDE's), variational distance uncertainty set models situations in which both the drift and diffusion coefficient of SDE's are unknown.

[1] The definition of total variation distance can be extended to signed measures.

[2] $\nu \in \mathscr{M}_1(\Sigma)$ is absolutely continuous with respect to $\mu \in \mathscr{M}_1(\Sigma)$, denoted by $\nu << \mu$, if $\mu(A) = 0$ for some $A \in \mathscr{B}(\Sigma)$ then $\nu(A) = 0$.

Define the spaces

$$BC(\Sigma) \stackrel{\triangle}{=} \{\ell : \Sigma \mapsto \mathbb{R} : \ell \text{ are bounded continuous}\},$$

$$BM(\Sigma) \stackrel{\triangle}{=} \{\ell : \Sigma \mapsto \mathbb{R} : \ell \text{ are bounded measurable functions}\},$$

$$BC^{+}(\Sigma) \stackrel{\triangle}{=} \{BC(\Sigma) : \ell \geq 0\}, \quad BM^{+}(\Sigma) \stackrel{\triangle}{=} \{BM(\Sigma) : \ell \geq 0\}.$$

$BC(\Sigma)$ and $BM(\Sigma)$ endowed with the sup norm $||\ell|| \stackrel{\triangle}{=} \sup_{x\in\Sigma} |\ell(x)|$, are Banach spaces [6]. Next, we introduce the two main extremum problems we shall investigate in this paper.

Problem II.1. *Given a fixed nominal distribution $\mu \in \mathscr{M}_1(\Sigma)$ and a parameter $R \in [0,2]$, define the class of true distributions by*

$$\mathbb{B}_R(\mu) \stackrel{\triangle}{=} \Big\{\nu \in \mathscr{M}_1(\Sigma) : ||\nu - \mu||_{TV} \leq R\Big\}, \tag{2}$$

and the average pay-off with respect to the true probability measure $\nu \in \mathbb{B}_R(\mu)$ by

$$\mathbb{L}_1(\nu) \stackrel{\triangle}{=} \int_\Sigma \ell(x)\nu(dx), \qquad \ell \in BC^{+}(\Sigma) \;\; or \;\; BM^{+}(\Sigma). \tag{3}$$

The objective is to find the extremum of the pay-off

$$D^{+}(R) \stackrel{\triangle}{=} \sup_{\nu\in\mathbb{B}_R(\mu)} \int_\Sigma \ell(x)\nu(dx). \tag{4}$$

Problem II.1 is a convex optimization problem on the space of probability measures. Note that, $BC^{+}(\Sigma)$, $BM^{+}(\Sigma)$ can be generalized to $L^{\infty,+}(\Sigma, \mathscr{B}(\Sigma), \nu)$, the set of all $\mathscr{B}(\Sigma)$-measurable, non-negative essentially bounded functions defined $\nu - a.e.$ endowed with the essential supremum norm $||\ell||_{\infty,\nu} = \nu\text{-ess }\sup_{x\in\Sigma} \ell(x) \stackrel{\triangle}{=} \inf_{\Delta\in\mathscr{N}_\eta} \sup_{x\in\Delta^c} \|\ell(x)\|$, where $\mathscr{N}_\nu = \{A \in \mathscr{B}(\Sigma) : \nu(A) = 0\}$.

In the context of minimax theory, Problem II.1 is important in uncertain stochastic control, estimation, and decision, formulated via minimax optimization. Such formulations are found in [7]–[11] utilizing relative entropy uncertainty, and in [12], [13] utilizing L_1 distance uncertainty. In the context of dynamic programming this is discussed in [14]. The second extremum problem is defined below.

Problem II.2. *Given a fixed nominal distribution $\mu \in \mathscr{M}_1(\Sigma)$ and a parameter $D \in [0,\infty)$, define the class of true distributions by*

$$\mathbb{Q}(D) \stackrel{\triangle}{=} \Big\{\nu \in \mathscr{M}_1(\Sigma) : \int_\Sigma \ell(x)\nu(dx) \leq D\Big\}, \qquad \ell \in BC^{+}(\Sigma) \;\; or \;\; BM^{+}(\Sigma), \tag{5}$$

and the total variation pay-off with respect to the true probability measure $\nu \in \mathbb{Q}(D)$ by

$$\mathbb{L}_2(\nu) \triangleq ||\nu - \mu||_{TV}. \tag{6}$$

The objective is to find the extremum of the pay-off

$$R^-(D) \triangleq \inf_{\nu \in \mathbb{Q}(D)} ||\nu - \mu||_{TV}, \tag{7}$$

whenever [3] $\int_\Sigma \ell(x)\mu(dx) > D$.

Problem II.2 is important in the context of approximation theory, since distance in variation is a measure of proximity of two probability distributions subject to constraints. It is also important in spectral measure or density approximation as follows. Recall that a function $\{R(\tau) : -\infty \leq \tau \leq \infty\}$ is the covariance function of a quadratic mean continuous and wide-sense stationary process if and only if it is of the form [15]

$$R(\tau) = \int_{-\infty}^{\infty} e^{2\pi\nu\tau} F(d\nu),$$

where $F(\cdot)$ is a finite Borel measure on $\mathbb{R}$, called spectral measure. Thus, by proper normalization of $F(\cdot)$ via $F_N(d\nu) \triangleq \frac{1}{R(0)} F(d\nu)$, then $F_N(d\nu)$ is a probability measure on $\mathscr{B}(\mathbb{R})$, and hence Problem II.2 can be used to approximate the class of spectral measures which satisfy moment estimates. Spectral estimation problems are discussed extensively in [16]–[20], utilizing relative entropy and Hellinger distances. However, in these references, the approximated spectral density is absolutely continuous with respect to the nominal spectral density; hence, it can not deal with reduced order approximation. In this respect, distance in total variation between spectral measures is very attractive.

A. Related Extremum Problems

Problems II.1, II.2 are related to additional extremum problems which are introduced below.

(1) The solution of (4) gives the solution to the problem defined by

$$R^+(D) \triangleq \sup_{\nu \in \mathscr{M}_1(\Sigma): \int_\Sigma \ell(x)\nu(dx) \leq D} ||\nu - \mu||_{TV}. \tag{8}$$

Specifically, $R^+(D)$ is the inverse mapping of $D^+(R)$. $D^+(R)$ is investigated in [21] in the context of minimax stochastic control under uncertainty, following an alternative approach which utilizes large deviation theory to express the extremum measure by a convex

[3] If $\int_\Sigma \ell(x)\mu(dx) \leq D$ then $\nu^* = \mu$ is the trivial extremum measure of (7).

combination of a tilted and the nominal probability measures. The two disadvantages of the method pursued in [8]–[11] are the following. 1) No explicit closed form expression for the extremum measure is given, and as a consequence, 2) its application to dynamic programming is restricted to a class of uncertain probability measures which are absolutely continuous with respect to the nominal measure $\mu(\Sigma) \in \mathscr{M}_1(\Sigma)$.

(2) Let ν and μ be absolutely continuous with respect to the Lebesgue measure so that $\varphi(x) \stackrel{\triangle}{=} \frac{d\nu}{d(x)}(x)$, $\psi(x) \stackrel{\triangle}{=} \frac{d\mu}{dx}(x)$ (e.g., $\varphi(\cdot)$, $\psi(\cdot)$ are the probability density functions of $\nu(\cdot)$ and $\mu(\cdot)$, respectively. Then, $||\nu-\mu||_{TV} = \int_\Sigma |\varphi(x)-\psi(x)|dx$ and hence, (4) and (8) are L_1-distance optimization problems.

(3) Let Σ be a non-empty denumerable set endowed with the discrete topology including finite cardinality $|\Sigma|$, with $\mathscr{M}_1(\Sigma)$ identified with the standard probability simplex in $\mathbb{R}^{|\Sigma|}$, that is, the set of all $|\Sigma|$-dimensional vectors which are probability vectors, and $\ell(x) \stackrel{\triangle}{=} -\log \nu(x), x\in\Sigma$, where $\{\nu(x) : x \in \Sigma\} \in \mathscr{M}_1(\Sigma)$, $\{\mu(x) : x \in \Sigma\} \in \mathscr{M}_1(\Sigma)$. Then (4) is equivalent to maximizing the entropy of $\{\nu(x) : x\in\Sigma\}$ subject to total variational distance metric constraint defined by

$$D^+(R) = \sup_{\nu\in\mathscr{M}_1(\Sigma):\sum_{x\in\Sigma}|\nu(x)-\mu(x)|\leq R} H(\nu). \tag{9}$$

Problem (9) is of interest when the concept of insufficient reasoning (e.g., Jayne's maximum entropy principle [22], [23]) is applied to construct a model for $\nu \in \mathscr{M}_1(\Sigma)$, subject to information quantified via total variational distance metric between ν and an empirical distribution μ. In the context of stochastic uncertain control systems, and its relation to robustness, Problem (9) with the total variational distance constraint replaced by relative entropy distance constraint is investigated in [24], [25].

(4) The solution of (7) gives the solution to the problem defined by

$$D^-(R) \stackrel{\triangle}{=} \inf_{\nu\in\mathscr{M}_1(\Sigma):||\nu-\mu||_{TV}\leq R} \int_\Sigma \ell(x)\nu(dx). \tag{10}$$

Problems (7) and (10) are important in approximating a class of probability distributions or spectral measures by reduced ones. In fact, the solution of (10) is obtained precisely as that of Problem II.1, with a reverse computation of the partition of the space Σ and the mass of the extremum measure on the partition moving in the opposite direction.

III. Characterization of Extremum Measures on Abstract Spaces

This section utilizes signed measures and some of their properties to convert Problems II.1, II.2 into equivalent extremum problems. First, we discuss some of the properties of these extremum Problems.

Lemma III.1.

(1) *$D^+(R)$ is a non-decreasing concave function of R, and*

$$D^+(R) = \sup_{||\nu-\mu||_{TV}=R} \int_\Sigma \ell(x)\nu(dx), \quad \text{if} \quad R \leq R_{\max}, \tag{11}$$

where $R_{\max}$ is the smallest non-negative number belonging to $[0,2]$ such that $D^+(R)$ is constant in $[R_{\max},2]$.

(2) *$R^-(D)$ is a non-increasing convex function of D, and*

$$R^-(D) = \inf_{\int_\Sigma \ell(x)\nu(dx)=D} ||\nu-\mu||_{TV}, \quad \text{if} \quad D \leq D_{\max}, \tag{12}$$

where $D_{\max}$ is the smallest non-negative number belonging to $[0,\infty)$ such that $R^-(D)=0$ for any $D \in [D_{\max},\infty)$.

Proof: **(1)** Suppose $0 \leq R_1 \leq R_2$, then for every $\nu \in \mathbb{B}_{R_1}(\mu)$ we have $||\nu-\mu||_{TV} \leq R_1 \leq R_2$, and therefore $\nu \in \mathbb{B}_{R_2}(\mu)$, hence

$$\sup_{\nu\in\mathbb{B}_{R_1}(\mu)} \int_\Sigma \ell(x)\nu(dx) \leq \sup_{\nu\in\mathbb{B}_{R_2}(\mu)} \int_\Sigma \ell(x)\nu(dx),$$

which is equivalent to $D^+(R_1) \leq D^+(R_2)$. So $D^+(R)$ is a non-decreasing function of R. Now consider two points $(R_1, D^+(R_1))$ and $(R_2, D^+(R_2))$ on the linear functional curve, such that $\nu_1 \in \mathbb{B}_{R_1}(\mu)$ achieves the supremum of (4) for R_1, and $\nu_2 \in \mathbb{B}_{R_2}(\mu)$ achieves the supremum of (4) for R_2. Then, $||\nu_1-\mu||_{TV} \leq R_1$ and $||\nu_2-\mu||_{TV} \leq R_2$. For any $\lambda \in (0,1)$, we have

$$||\lambda\nu_1+(1-\lambda)\nu_2-\mu||_{TV} \leq \lambda||\nu_1-\mu||_{TV}+(1-\lambda)||\nu_2-\mu||_{TV} \leq \lambda R_1+(1-\lambda)R_2 = R.$$

Define $\nu^* \triangleq \lambda\nu_1+(1-\lambda)\nu_2$, $R \triangleq \lambda R_1+(1-\lambda)R_2$. The previous equation implies that $\nu^* \in \mathbb{B}_R(\mu)$, hence $D^+(\lambda R_1+(1-\lambda)R_2) \geq \int_\Sigma \ell(x)\nu^*(dx)$. Therefore,

$$\begin{aligned} D^+(R) &= \sup_{\nu\in\mathbb{B}_R(\mu)} \int_\Sigma \ell(x)\nu(dx) \geq \int_\Sigma \ell(x)\nu^*(dx) = \int_\Sigma \ell(x)\left(\lambda\nu_1(dx)+(1-\lambda)\nu_2(dx)\right) \\ &= \lambda\int_\Sigma \ell(x)\nu_1(dx)+(1-\lambda)\int_\Sigma \ell(x)\nu_2(dx) = \lambda D^+(R_1)+(1-\lambda)D^+(R_2). \end{aligned}$$

So, $D^+(R)$ is a concave function of R. Also the right side of (11), say $\bar{D}^+(R)$, is concave function of R. But $D^+(R) = \sup_{R' \leq R} \bar{D}^+(R')$ which completes the derivation of (11).

(2) Suppose $0 \leq D_1 \leq D_2$, then $\mathbb{Q}(D_1) \subset \mathbb{Q}(D_2)$, and $\inf_{\nu \in \mathbb{Q}(D_1)} ||\nu - \mu||_{TV} \geq \inf_{\nu \in \mathbb{Q}(D_2)} ||\nu - \mu||_{TV}$ which is equivalent to $R^-(D_1) \geq R^-(D_2)$. Hence, $R^-(D)$ is a non-increasing function of D. Now consider two points $(D_1, R^-(D_1))$ and $(D_2, R^-(D_2))$ on the total variation curve. Let $D \triangleq \lambda D_1 + (1-\lambda) D_2$, $\nu^* \triangleq \lambda \nu_1 + (1-\lambda)\nu_2$ and $\nu_1 \in \mathbb{Q}(D_1)$, $\nu_2 \in \mathbb{Q}(D_2)$ such that $||\nu_1 - \mu||_{TV} = R^-(D_1)$ and $||\nu_2 - \mu||_{TV} = R^-(D_2)$. Then, $\int_\Sigma \ell(x)\nu_1(dx) \leq D_1$ and $\int_\Sigma \ell(x)\nu_2(dx) \leq D_2$. Taking convex combination leads to

$$\lambda \int_\Sigma \ell(x)\nu_1(dx) + (1-\lambda)\int_\Sigma \ell(x)\nu_2(dx) \leq \lambda D_1 + (1-\lambda)D_2 = D,$$

and hence $\nu^* \in \mathbb{Q}(D)$. So,

$$\begin{aligned} R^-(D) &= \inf_{\nu \in \mathbb{Q}(D)} ||\nu - \mu||_{TV} \leq ||\nu^* - \mu||_{TV} = ||\lambda \nu_1 + (1-\lambda)\nu_2 - \mu||_{TV} \\ &\leq \lambda ||\nu_1 - \mu||_{TV} + (1-\lambda)||\nu_2 - \mu||_{TV} = \lambda R^-(D_1) + (1-\lambda) R^-(D_2). \end{aligned}$$

This shows that $R^-(D)$ is convex function of D. Also the right side of (12), say $\bar{R}^-(D)$, is convex function of D. But, $R^-(D) = \inf_{D' \leq D} \bar{R}^-(D')$ which completes the derivation of (12). ■

Let $\mathcal{M}_{sm}(\Sigma)$ denote the set of finite signed measures. Then, any $\eta \in \mathcal{M}_{sm}(\Sigma)$ has a Jordan decomposition [26] $\{\eta^+, \eta^-\}$ such that $\eta = \eta^+ - \eta^-$, and the total variation of η is defined by $||\eta||_{TV} \triangleq \eta^+(\Sigma) + \eta^-(\Sigma)$. Define the following subset $\mathbb{M}_0(\Sigma) \triangleq \Big\{\eta \in \mathcal{M}_{sm}(\Sigma) : \eta(\Sigma) = 0\Big\}$. For $\xi \in \mathbb{M}_0(\Sigma)$, then $\xi(\Sigma) = 0$, which implies that $\xi^+(\Sigma) = \xi^-(\Sigma)$, and hence $\xi^+(\Sigma) = \xi^-(\Sigma) = \frac{||\xi||_{TV}}{2}$. Then, $\xi \triangleq \nu - \mu \in \mathbb{M}_0(\Sigma)$ and hence $\xi = (\nu-\mu)^+ - (\nu-\mu)^- \equiv \xi^+ - \xi^-$.

A. *Equivalent Extremum Problem of $D^+(R)$*

Consider the pay-off of Problem II.1, for $\ell \in BC^+(\Sigma)$. Then the following inequalities hold.

$$\begin{aligned} D^+(R) \triangleq \int_\Sigma \ell(x)\nu(dx) &\overset{(a)}{=} \int_\Sigma \ell(x)\left(\xi^+(dx) - \xi^-(dx)\right) + \int_\Sigma \ell(x)\mu(dx) \\ &\overset{(b)}{\leq} \sup_{x\in\Sigma} \ell(x)\xi^+(\Sigma) - \inf_{x\in\Sigma} \ell(x)\xi^-(\Sigma) + \int_\Sigma \ell(x)\mu(dx) \\ &\overset{(c)}{=} \sup_{x\in\Sigma} \ell(x)\frac{||\xi||_{TV}}{2} - \inf_{x\in\Sigma} \ell(x)\frac{||\xi||_{TV}}{2} + \int_\Sigma \ell(x)\mu(dx) \\ &= \left\{\sup_{x\in\Sigma} \ell(x) - \inf_{x\in\Sigma} \ell(x)\right\}\frac{||\xi||_{TV}}{2} + \int_\Sigma \ell(x)\mu(dx), \end{aligned} \tag{13}$$

where (a) follows by adding and subtracting $\int \ell d\mu$, and from the Jordan decomposition of $(\nu - \mu)$, (b) follows due to $\ell \in BC^+(\Sigma)$, (c) follows because any $\xi \in \mathbb{M}_0(\Sigma)$ satisfies $\xi^+(\Sigma) = \xi^-(\Sigma) = \frac{1}{2}||\xi||_{TV}$. For a given $\mu \in \mathscr{M}_1(\Sigma)$ and $\nu \in \mathbb{B}_R(\mu)$ define the set

$$\widetilde{\mathbb{B}}_R(\mu) \triangleq \{\xi \in \mathbb{M}_0(\Sigma) : \xi = \nu - \mu, \nu \in \mathscr{M}_1(\Sigma), ||\xi||_{TV} \leq R\}.$$

The upper bound in the right hand side of (13) is achieved by $\xi^* \in \widetilde{\mathbb{B}}_R(\mu)$ as follows. Let

$$x^0 \in \Sigma^0 \triangleq \{x \in \Sigma : \ell(x) = \sup\{\ell(x) : x \in \Sigma\} \equiv M\},$$
$$x_0 \in \Sigma_0 \triangleq \{x \in \Sigma : \ell(x) = \inf\{\ell(x) : x \in \Sigma\} \equiv m\}.$$

Take

$$\xi^*(dx) = \nu^*(dx) - \mu(dx) = \frac{R}{2}\left(\delta_{x^0}(dx) - \delta_{x_0}(dx)\right), \tag{14}$$

where $\delta_y(dx)$ denotes the Dirac measure concentrated at $y \in \Sigma$. This is indeed a signed measure with total variation $||\xi^*||_{TV} = ||\nu^* - \mu||_{TV} = R$, and $\int_\Sigma \ell(x)(\nu^* - \mu)(dx) = \frac{R}{2}(M - m)$. Hence, by using (14) as a candidate of the maximizing distribution then the extremum Problem II.1 is equivalent to

$$D^+(R) = \int_\Sigma \ell(x)\nu^*(dx) = \frac{R}{2}\left\{\sup_{x\in\Sigma}\ell(x) - \inf_{x\in\Sigma}\ell(x)\right\} + E_\mu(\ell), \tag{15}$$

where ν^* satisfies the constraint $||\xi^*||_{TV} = ||\nu^* - \mu||_{TV} = R$, it is normalized $\nu^*(\Sigma) = 1$, and $0 \leq \nu^*(A) \leq 1$ on any $A \in \mathscr{B}(\Sigma)$. Alternatively, the pay-off $\int_\Sigma \ell(x)\nu^*(dx)$ can be written as

$$\int_\Sigma \ell(x)\nu^*(dx) = \int_{\Sigma^0} M\nu^*(dx) + \int_{\Sigma_0} m\nu^*(dx) + \int_{\Sigma\setminus\Sigma^0\cup\Sigma_0} \ell(x)\mu(dx). \tag{16}$$

Hence, the optimal distribution $\nu^* \in \mathbb{B}_R(\mu)$ satisfies

$$\begin{aligned}&\int_{\Sigma^0}\nu^*(dx) = \mu(\Sigma^0) + \frac{R}{2} \in [0,1], \quad \int_{\Sigma_0}\nu^*(dx) = \mu(\Sigma_0) - \frac{R}{2} \in [0,1],\\ &\nu^*(A) = \mu(A), \quad \forall A \subseteq \Sigma\setminus\Sigma^0\cup\Sigma_0.\end{aligned} \tag{17}$$

Remark III.2.

(1) *For $\mu \in \mathscr{M}_1(\Sigma)$ which do not include point mass, and for $f \in BC^+(\Sigma)$, if Σ^0 and Σ_0 are countable, then (17) is $\mu(\Sigma^0) = \mu(\Sigma_0) = 0$, $\nu^*(\Sigma_0) = 0$, $\nu^*(\Sigma^0) = \frac{R}{2}$, $\nu^*(\Sigma\setminus\Sigma^0\cup\Sigma_0) = \mu(\Sigma\setminus\Sigma^0\cup\Sigma_0) - \frac{R}{2}$.*

(2) *The first right side term in (15) is related to the oscillator seminorm of $f \in BM(\Sigma)$ called global modulus of continuity, defined by $\operatorname{osc}(f) \triangleq \sup_{(x,y)\in\Sigma\times\Sigma}|f(x) - f(y)| = 2\inf_{\alpha\in\mathbb{R}}||f - \alpha||$. For $f \in BM^+(\Sigma)$, $\operatorname{osc}(f) = \sup_{x\in\Sigma}|f(x)| - \inf_{x\in\Sigma}|f(x)|$.*

B. Equivalent Extremum Problem of $R^-(D)$

Next, we proceed with the abstract formulation of Problem II.2. Consider the constraint of Problem II.2, for $\ell \in BC^+(\Sigma)$. Then the following inequalities hold.

$$\begin{aligned}\int_\Sigma \ell(x)\nu(dx) &= \int_\Sigma \ell(x)\left(\xi^+(dx)-\xi^-(dx)\right)+\int_\Sigma \ell(x)\mu(dx)\\ &\geq \inf_{x\in\Sigma}\ell(x)\xi^+(\Sigma)-\sup_{x\in\Sigma}\ell(x)\xi^-(\Sigma)+\int_\Sigma \ell(x)\mu(dx)\\ &= \inf_{x\in\Sigma}\ell(x)\frac{||\xi||_{TV}}{2}-\sup_{x\in\Sigma}\ell(x)\frac{||\xi||_{TV}}{2}+\int_\Sigma \ell(x)\mu(dx)\\ &= \left\{\inf_{x\in\Sigma}\ell(x)-\sup_{x\in\Sigma}\ell(x)\right\}\frac{||\xi||_{TV}}{2}+\int_\Sigma \ell(x)\mu(dx). \end{aligned} \tag{18}$$

The lower bound on the right hand side of (18) is achieved by choosing $\xi^* \in \widetilde{\mathbb{B}}_R(\mu)$ as follows

$$\xi^*(dx) = \nu^*(dx)-\mu(dx) = \frac{R}{2}\left(\delta_{x_0}(dx)-\delta_{x^0}(dx)\right). \tag{19}$$

This is a signed measure with total variation $||\xi^*||_{TV} = ||\nu^*-\mu||_{TV} = R$. Hence, by using (19) as a candidate of the minimizing distribution then (18) is equivalent to

$$\int_\Sigma \ell(x)\nu^*(dx) = \frac{R}{2}\left\{\inf_{x\in\Sigma}\ell(x)-\sup_{x\in\Sigma}\ell(x)\right\}+E_\mu(\ell). \tag{20}$$

Solving the above equation with respect to R the extremum Problem II.2 (for $D < E_\mu(\ell)$) is equivalent to

$$R^-(D) = \frac{2(D-E_\mu(\ell))}{\left\{\inf_{x\in\Sigma}\ell(x)-\sup_{x\in\Sigma}\ell(x)\right\}}, \tag{21}$$

where ν^* satisfies the constraint $\int_\Sigma \ell(x)\nu^*(dx) = D$, it is normalized $\nu^*(\Sigma)=1$, and $0\leq \nu(A)\leq 1$ on any $A\in\mathscr{B}(\Sigma)$. We can now identify $R_{\max}$ and $D_{\max}$ described in Lemma III.1. These are stated as a corollary.

Corollary III.3. *The values of $R_{\max}$ and $D_{\max}$ described in Lemma III.1 are given by*

$$R_{\max} = 2\left(1-\mu\left(\Sigma^0\right)\right) \quad \text{and} \quad D_{\max} = \int_\Sigma \ell(x)\mu(dx).$$

Proof: Concerning $R_{\max}$, we know that $D^+(R) \leq \sup_{x\in\Sigma}\ell(x)$, $\forall R\geq 0$, hence $D^+(R_{\max})$ can be at most $\sup_{x\in\Sigma}\ell(x)$. Since $D^+(R)$ is non-decreasing then $D^+(R_{\max}) \leq D^+(R) \leq \sup_{x\in\Sigma}\ell(x)$, for any $R\geq R_{\max}$. Consider a ν that achieves this supremum. Let $\mu(\Sigma^0)$ and $\nu(\Sigma^0)$ to denote

the nominal and true probability measures on Σ^0, respectively. If $\nu(\Sigma^0)=1$ then $\nu(\Sigma\setminus\Sigma^0)=0$. Therefore,

$$\begin{aligned}||\nu-\mu||_{TV} &= \sum_{x\in\Sigma^0}|\nu(x)-\mu(x)|+\sum_{x\in\Sigma\setminus\Sigma^0}|\nu(x)-\mu(x)| \overset{(a)}{=} \sum_{x\in\Sigma^0}|\nu(x)-\mu(x)|+\sum_{x\in\Sigma\setminus\Sigma^0}|-\mu(x)| \\ &\overset{(b)}{=} \sum_{x\in\Sigma^0}\nu(x)-\sum_{x\in\Sigma^0}\mu(x)+\sum_{x\in\Sigma\setminus\Sigma^0}\mu(x)=1-\sum_{x\in\Sigma^0}\mu(x)+\sum_{x\in\Sigma\setminus\Sigma^0}\mu(x) \\ &= 2\left(1-\sum_{x\in\Sigma^0}\mu(x)\right)=2\left(1-\mu(\Sigma^0)\right),\end{aligned}$$

where (a) follows due to $\nu(\Sigma\setminus\Sigma^0)=0$ which implies $\nu(x)=0$ for any $x\in\Sigma\setminus\Sigma^0$, and (b) follows because $\nu(x)\geq\mu(x)$ for all $x\in\Sigma^0$. Therefore, $R_{\max}=2(1-\mu(\Sigma^0))$ implies that $D^+(R_{\max})=\sup_{x\in\Sigma}\ell(x)$. Hence, $D^+(R)=\sup_{x\in\Sigma}\ell(x)$, for any $R\geq R_{\max}$.

Concerning $D_{\max}$, we know that $R^-(D)\geq 0$ for all $D\geq 0$ hence $R^-(D_{\max})$ can be at least zero. Let $D_{\max}=\int_\Sigma \ell(x)\mu(dx)$, then it is obvious that $R^-(D_{\max})=0$. Since $R^-(D)$ in non-increasing, then $0\leq R^-(D)\leq R^-(D_{\max})$, for any $D\geq D_{\max}$. Hence, $R^-(D)=0$, for any $D\geq D_{\max}$. ■

IV. Characterization of Extremum Measures for Finite Alphabets

This section uses the results of Section III to compute closed form expressions for the extremum measures ν^* for any $R\in[0,2]$, when Σ is a finite alphabet space to give the intuition into the solution procedure. This is done by identifying the sets Σ^0, Σ_0, $\Sigma\setminus\Sigma^0\cup\Sigma_0$, and the measure ν^* on these sets for any $R\in[0,2]$. Although this can be done for probability measures on complete separable metric spaces (Polish spaces) (Σ,d_Σ), and for $\ell\in BM^+(\Sigma)$, $\ell\in BC^+(\Sigma)$, $L^{\infty,+}(\Sigma,\mathscr{B}(\Sigma),\nu)$, we prefer to discuss the finite alphabet case to gain additional insight into these problems. At the end of this section we shall use the finite alphabet case to discuss the extensions to countable alphabet and to $\ell\in L^{\infty,+}(\Sigma,\mathscr{B}(\Sigma),\nu)$.

Consider the finite alphabet case $(\Sigma,\mathscr{M})$, where $\text{card}(\Sigma)=|\Sigma|$ is finite, $\mathscr{M}=2^{|\Sigma|}$. Thus, ν and μ are point mass distributions on Σ. Define the set of probability vectors on Σ by

$$\mathbb{P}(\Sigma)\triangleq\Big\{p=(p_1,\ldots,p_{|\Sigma|}):p_i\geq 0, i=0,\ldots,|\Sigma|,\sum_{i\in\Sigma}p_i=1\Big\}. \tag{22}$$

Thus, $p\in\mathbb{P}(\Sigma)$ is a probability vector in $\mathbb{R}_+^{|\Sigma|}$. Also let $\ell\triangleq\{\ell_1,\ldots,\ell_{|\Sigma|}\}$ so that $\ell\in\mathbb{R}_+^{|\Sigma|}$ (e.g., set of non-negative vectors of dimension $|\Sigma|$).

A. *Problem II.1: Finite Alphabet Case*

Suppose $\nu \in \mathbb{P}(\Sigma)$ is the true probability vector and $\mu \in \mathbb{P}(\Sigma)$ is the nominal fixed probability vector. The extremum problem is defined by

$$D^+(R) \stackrel{\triangle}{=} \max_{\nu \in \mathbb{B}_R(\mu)} \sum_{i \in \Sigma} \ell_i \nu_i, \tag{23}$$

where

$$\mathbb{B}_R(\mu) \stackrel{\triangle}{=} \Big\{ \nu \in \mathbb{P}(\Sigma)) : ||\nu - \mu||_{TV} \triangleq \sum_{i \in \Sigma} |\nu_i - \mu_i| \leq R \Big\}. \tag{24}$$

Next, we apply the results of Section III to characterize the optimal ν^* for any $R \in [0,2]$. By defining, $\xi_i \stackrel{\triangle}{=} \nu_i - \mu_i, i = 1, \ldots, |\Sigma|$ and $\xi \in \mathbb{M}_0(\Sigma)$, Problem II.1 can be reformulated as follows.

$$\max_{\nu \in \mathbb{B}_R(\mu)} \sum_{i \in \Sigma} \ell_i \nu_i \longrightarrow \sum_{i \in \Sigma} \ell_i \mu_i + \max_{\xi \in \widetilde{\mathbb{B}}_R(\mu)} \sum_{i \in \Sigma} \ell_i \xi_i. \tag{25}$$

Note that $\xi \in \widetilde{\mathbb{B}}_R(\mu)$ is described by the constraints

$$\sum_{i \in \Sigma} |\xi_i| \leq R, \quad \sum_{i \in \Sigma} \xi_i = 0, \quad 0 \leq \xi_i + \mu_i \leq 1, \quad \forall i \in \Sigma. \tag{26}$$

The positive and negative variation of the signed measure ξ are defined by

$$\xi_i^+ \triangleq \begin{cases} \xi_i, & \text{if } \xi_i \geq 0 \\ 0, & \text{if } \xi_i < 0, \end{cases} \qquad \xi_i^- \triangleq \begin{cases} 0, & \text{if } \xi_i \geq 0 \\ -\xi_i, & \text{if } \xi_i < 0, \end{cases}$$

Therefore,

$$\sum_{i \in \Sigma} \xi_i = \sum_{i \in \Sigma} \xi_i^+ - \sum_{i \in \Sigma} \xi_i^-, \qquad \sum_{i \in \Sigma} |\xi_i| = \sum_{i \in \Sigma} \xi_i^+ + \sum_{i \in \Sigma} \xi_i^-,$$

and hence,

$$\sum_{i \in \Sigma} \xi_i^+ = \frac{\sum_{i \in \Sigma} \xi_i + \sum_{i \in \Sigma} |\xi_i|}{2} \equiv \frac{\alpha}{2}, \qquad \sum_{i \in \Sigma} \xi_i^- = \frac{-\sum_{i \in \Sigma} \xi_i + \sum_{i \in \Sigma} |\xi_i|}{2} \equiv \frac{\alpha}{2},$$

and

$$\sum_{i \in \Sigma} \xi_i = 0, \qquad \alpha = \sum_{i \in \Sigma} |\xi_i| \leq R. \tag{27}$$

In addition,

$$\sum_{i \in \Sigma} \ell_i \xi_i = \sum_{i \in \Sigma} \ell_i \xi_i^+ - \sum_{i \in \Sigma} \ell_i \xi_i^-. \tag{28}$$

Define the maximum and minimum values of the sequence $\{\ell_1, \ldots, \ell_{|\Sigma|}\}$ by $\ell_{\max} \triangleq \max_{i \in \Sigma} \ell_i$, $\ell_{\min} \triangleq \min_{i \in \Sigma} \ell_i$, and its corresponding support sets by $\Sigma^0 \triangleq \{i \in \Sigma : \ell_i = \ell_{\max}\}$, $\Sigma_0 \triangleq \{i \in \Sigma :$

$\ell_i = \ell_{\min}\}$. For all remaining sequence, $\{\ell_i : i \in \Sigma \setminus \Sigma^0 \cup \Sigma_0\}$, and for $1 \leq r \leq |\Sigma \setminus \Sigma^0 \cup \Sigma_0|$ define recursively

$$\Sigma_k \triangleq \left\{ i \in \Sigma : \ell_i = \min\left\{ \ell_\alpha : \alpha \in \Sigma \setminus \Sigma^0 \cup \left(\bigcup_{j=1}^{k} \Sigma_{j-1} \right) \right\} \right\}, \quad k \in \{1, 2, \ldots, r\}, \tag{29}$$

till all the elements of Σ are exhausted (i.e., k is at most $|\Sigma \setminus \Sigma^0 \cup \Sigma_0|$). Define the corresponding values of the sequence of sets in (29) by

$$\ell(\Sigma_k) \triangleq \min_{i \in \Sigma \setminus \Sigma^0 \cup (\bigcup_{j=1}^{k} \Sigma_{j-1})} \ell_i, \quad k \in \{1, 2, \ldots, r\},$$

where r is the number of Σ_k sets which is at most $|\Sigma \setminus \Sigma^0 \cup \Sigma_0|$; for example, when $k = 1$, $\ell(\Sigma_1) = \min_{i \in \Sigma \setminus \Sigma^0 \cup \Sigma_0} \ell_i$. The following theorem characterizes the solution of Problem II.1.

Theorem IV.1. *The solution of the finite alphabet version of Problem II.1 is given by*

$$D^+(R) = \ell_{\max} \nu^*(\Sigma^0) + \ell_{\min} \nu^*(\Sigma_0) + \sum_{k=1}^{r} \ell(\Sigma_k) \nu^*(\Sigma_k). \tag{30}$$

Moreover, the optimal probabilities are given by

$$\nu^*(\Sigma^0) \triangleq \sum_{i \in \Sigma^0} \nu_i^* = \sum_{i \in \Sigma^0} \mu_i + \alpha, \tag{31a}$$

$$\nu^*(\Sigma_0) \triangleq \sum_{i \in \Sigma_0} \nu_i^* = \left(\sum_{i \in \Sigma_0} \mu_i - \alpha \right)^+, \tag{31b}$$

$$\nu^*(\Sigma_k) \triangleq \sum_{i \in \Sigma_k} \nu_i^* = \left(\sum_{i \in \Sigma_k} \mu_i - \left(\alpha - \sum_{j=1}^{k} \sum_{i \in \Sigma_{j-1}} \mu_i \right)^+ \right)^+, \tag{31c}$$

$$\alpha = \min\left(\frac{R}{2}, 1 - \sum_{i \in \Sigma^0} \mu_i \right), \tag{31d}$$

where, $k = 1, 2, \ldots, r$ and r is the number of Σ_k sets which is at most $|\Sigma \setminus \Sigma^0 \cup \Sigma_0|$.

Proof: The derivation of the Theorem is based on a sequence of Lemmas, Propositions and Corollaries which are presented below. ■

The following Lemma is a direct consequence of Section III-A.

Lemma IV.2. *Consider the finite alphabet version of Problem II.1. Then the following bounds hold.*

1. Upper Bound.

$$\sum_{i\in\Sigma}\ell_i\xi_i^+ \leq \ell_{\max}\left(\frac{\alpha}{2}\right). \tag{32}$$

The upper bound holds with equality if

$$\sum_{i\in\Sigma^0}\mu_i+\frac{\alpha}{2}\leq 1,\quad \sum_{i\in\Sigma^0}\xi_i^+=\frac{\alpha}{2},\quad \xi_i^+=0 \;\; \textit{for} \;\; i\in\Sigma\setminus\Sigma^0, \tag{33}$$

and the optimal probability on Σ^0 *is given by*

$$\nu^*(\Sigma^0)\triangleq\sum_{i\in\Sigma^0}\nu_i^*=\min\left(1,\sum_{i\in\Sigma^0}\mu_i+\frac{\alpha}{2}\right). \tag{34}$$

2. Lower Bound.

$$\sum_{i\in\Sigma}\ell_i\xi_i^- \geq \ell_{\min}\left(\frac{\alpha}{2}\right). \tag{35}$$

The lower bound holds with equality if

$$\sum_{i\in\Sigma_0}\mu_i-\frac{\alpha}{2}\geq 0,\quad \sum_{i\in\Sigma_0}\xi_i^-=\frac{\alpha}{2},\quad \xi_i^-=0 \;\; \textit{for} \;\; i\in\Sigma\setminus\Sigma_0, \tag{36}$$

and the optimal probability on Σ_0 *is given by*

$$\nu^*(\Sigma_0)\triangleq\sum_{i\in\Sigma_0}\nu_i^*=\left(\sum_{i\in\Sigma_0}\mu_i-\frac{\alpha}{2}\right)^+. \tag{37}$$

Moreover, under the conditions in 1 and 2 the maximum pay-off is given by

$$D^+(R)=\frac{\alpha}{2}\{\ell_{\max}-\ell_{\min}\}+\sum_{i\in\Sigma}\ell_i\mu_i. \tag{38}$$

Proof: Follows from Section III-A. ■

Proposition IV.3. *If* $\sum_{i\in\Sigma^0}\mu_i+\frac{\alpha}{2}=1$ *then* $D^+(R)=\ell_{\max}$.

Proof: Under the stated condition $\sum_{i\in\Sigma^0}\nu_i^*=1$ and therefore $\sum_{i\in\Sigma\setminus\Sigma^0}\nu_i^*=0$, hence $\nu_i^*=0$, for all $i\in\Sigma\setminus\Sigma^0$. Then the maximum pay-off (23) is given by

$$D^+(R)=\sum_{i\in\Sigma^0}\ell_i\nu_i^*+\sum_{i\in\Sigma\setminus\Sigma^0}\ell_i\nu_i^*=\ell_{\max}\sum_{i\in\Sigma^0}\nu_i^*=\ell_{\max}.$$

■

The lower bound of Lemma IV.2 characterize the extremum solution for $\sum_{i\in\Sigma_0}\mu_i-\frac{\alpha}{2}\geq 0$. Next, the characterization of extremum solution is discussed when this condition is violated.

Lemma IV.4. *If* $\sum_{i\in\Sigma_0}\mu_i - \frac{\alpha}{2} \leq 0$, *then*

$$\sum_{i\in\Sigma} \ell_i \xi_i^- \geq \ell(\Sigma_1)\left(\frac{\alpha}{2} - \sum_{i\in\Sigma_0}\mu_i\right) + \ell_{\min}\sum_{i\in\Sigma_0}\mu_i. \tag{39}$$

Moreover, equality holds if

$$\sum_{i\in\Sigma_0}\xi_i^- = \sum_{i\in\Sigma_0}\mu_i, \tag{40a}$$

$$\sum_{i\in\Sigma_1}\xi_i^- = \left(\frac{\alpha}{2} - \sum_{i\in\Sigma_0}\mu_i\right), \tag{40b}$$

$$\sum_{i\in\Sigma_0}\mu_i + \sum_{i\in\Sigma_1}\mu_1 \geq \frac{\alpha}{2}, \tag{40c}$$

$$\xi_i^- = 0 \ \text{ for all } \ i \in \Sigma \setminus \Sigma_0 \cup \Sigma_1, \tag{40d}$$

and the optimal probability on Σ_1 *is given by*

$$\sum_{i\in\Sigma_1}\nu_i^* = \left(\sum_{i\in\Sigma_1}\mu_i - \left(\frac{\alpha}{2} - \sum_{i\in\Sigma_0}\mu_i\right)^+\right)^+. \tag{41}$$

Proof: First, we show that inequality holds.

$$\sum_{i\in\Sigma\setminus\Sigma_0}\ell_i\xi_i^- \geq \min_{i\in\Sigma\setminus\Sigma_0}\ell_i \sum_{i\in\Sigma\setminus\Sigma_0}\xi_i^- = \ell(\Sigma_1)\sum_{i\in\Sigma\setminus\Sigma_0}\xi_i^- = \ell(\Sigma_1)\left(\sum_{i\in\Sigma}\xi_i^- - \sum_{i\in\Sigma_0}\xi_i^-\right).$$

Hence,

$$\sum_{i\in\Sigma}\ell_i\xi_i^- - \sum_{i\in\Sigma_0}\ell_i\xi_i^- \geq \ell(\Sigma_1)\left(\frac{\alpha}{2} - \sum_{i\in\Sigma_0}\mu_i\right),$$

which implies

$$\sum_{i\in\Sigma}\ell_i\xi_i^- \geq \ell(\Sigma_1)\left(\frac{\alpha}{2} - \sum_{i\in\Sigma_0}\mu_i\right) + \ell_{\min}\sum_{i\in\Sigma_0}\mu_i,$$

establishing (39). Next, we show under the stated conditions that equality holds.

$$\begin{aligned}\sum_{i\in\Sigma}\ell_i\xi_i^- &= \sum_{i\in\Sigma_0}\ell_i\xi_i^- + \sum_{i\in\Sigma_1}\ell_i\xi_i^- + \sum_{i\in\Sigma\setminus\Sigma_0\cup\Sigma_1}\ell_i\xi_i^- \\ &= \ell_{\min}\sum_{i\in\Sigma_0}\mu_i + \ell(\Sigma_1)\sum_{i\in\Sigma_1}\xi_i^- = \ell_{\min}\sum_{i\in\Sigma_0}\mu_i + \ell(\Sigma_1)\left(\frac{\alpha}{2} - \sum_{i\in\Sigma_0}\mu_i\right).\end{aligned}$$

From (40b) we have that

$$\sum_{i\in\Sigma_1}\xi_i^- = \left(\frac{\alpha}{2} - \sum_{i\in\Sigma_0}\mu_i\right), \tag{42}$$

and hence,

$$\sum_{i\in\Sigma_1} v_i = \sum_{i\in\Sigma_1}\mu_i - \left(\frac{\alpha}{2} - \sum_{i\in\Sigma_0}\mu_i\right). \tag{43}$$

The optimal $\sum_{i\in\Sigma_1} v_i$ must satisfy $\frac{\alpha}{2} - \sum_{i\in\Sigma_0}\mu_i \geq 0$ and $\sum_{i\in\Sigma_1}\mu_i + \sum_{i\in\Sigma_0}\mu_i - \frac{\alpha}{2} \geq 0$. Hence, (41) is obtained. ■

Following the previous Lemma, which characterizes the extremum solution when $\sum_{i\in\Sigma_0}\mu_i - \frac{\alpha}{2} \leq 0$, one can also characterize the optimum solution of extremum Problem II.1, when $\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i - \frac{\alpha}{2} \leq 0$, for any $k \in \{1,2,\ldots,r\}$.

Corollary IV.5. *For any $k \in \{1,2,\ldots,r\}$, if $\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i - \frac{\alpha}{2} \leq 0$ then*

$$\sum_{i\in\Sigma}\ell_i\xi_i^- \geq \ell(\Sigma_k)\left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right) + \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\ell_i\mu_i. \tag{44}$$

Moreover, equality holds if

$$\sum_{i\in\Sigma_{j-1}}\xi_i^- = \sum_{i\in\Sigma_{j-1}}\mu_i, \quad \text{for all} \quad j = 1,2,\ldots,k, \tag{45a}$$

$$\sum_{i\in\Sigma_k}\xi_i^- = \left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right), \tag{45b}$$

$$\sum_{j=0}^{k}\sum_{i\in\Sigma_j}\mu_i \geq \frac{\alpha}{2}, \tag{45c}$$

$$\xi_i^- = 0 \quad \text{for all} \quad i \in \Sigma\setminus\Sigma_0\cup\Sigma_1\cup\ldots\cup\Sigma_k, \tag{45d}$$

and the optimal probability on Σ_k sets is given by

$$\sum_{i\in\Sigma_k} v_i^* = \left(\sum_{i\in\Sigma_k}\mu_i - \left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right)^+\right)^+. \tag{46}$$

Proof: Consider any $k \in \{1,2,\ldots,r\}$. First, we show that inequality holds. From lower bound we have that

$$\begin{aligned}\sum_{i\in\Sigma\setminus\cup_{j=1}^{k}\Sigma_{j-1}}\ell_i\xi_i^- &\geq \min_{i\in\Sigma\setminus\cup_{j=1}^{k}\Sigma_{j-1}}\ell_i\sum_{i\in\Sigma\setminus\cup_{j=1}^{k}\Sigma_{j-1}}\xi_i^- \\ &= \ell(\Sigma_k)\sum_{i\in\Sigma\setminus\cup_{j=1}^{k}\Sigma_{j-1}}\xi_i^- = \ell(\Sigma_k)\left(\sum_{i\in\Sigma}\xi_i^- - \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\xi_i^-\right).\end{aligned}$$

Hence,

$$\sum_{i\in\Sigma}\ell_i\xi_i^- - \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\ell_i\xi_i^- \geq \ell(\Sigma_k)\left(\frac{\alpha}{2}-\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right),$$

which implies

$$\sum_{i\in\Sigma}\ell_i\xi_i^- \geq \ell(\Sigma_k)\left(\frac{\alpha}{2}-\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right)+\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\ell_i\mu_i.$$

Next, we show under the stated conditions that equality holds.

$$\begin{aligned}\sum_{i\in\Sigma}\ell_i\xi_i^- &= \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\ell_i\xi_i^- + \sum_{i\in\Sigma_k}\ell_i\xi_i^- + \sum_{i\in\Sigma\setminus\cup_{j=0}^{k}\Sigma_j}\ell_i\xi_i^- \\ &= \sum_{j=1}^{k}\ell(\Sigma_{j-1})\sum_{i\in\Sigma_{j-1}}\xi_i^- + \ell(\Sigma_k)\sum_{i\in\Sigma_k}\xi_i^- = \sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\ell_i\mu_i + \ell(\Sigma_k)\left(\frac{\alpha}{2}-\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right).\end{aligned}$$

From (45b) we have that

$$\sum_{i\in\Sigma_k}\xi_i^- = \left(\frac{\alpha}{2}-\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right), \tag{47}$$

and hence,

$$\sum_{i\in\Sigma_k}\nu_i = \sum_{i\in\Sigma_k}\mu_i - \left(\frac{\alpha}{2}-\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i\right). \tag{48}$$

The optimal $\sum_{i\in\Sigma_k}\nu_i^*$ must satisfy $\frac{\alpha}{2}-\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i \geq 0$ and $\sum_{j=0}^{k}\sum_{i\in\Sigma_j}\mu_i - \frac{\alpha}{2} \geq 0$. Hence, (46) is obtained. ■

Putting together Lemma IV.2, Proposition IV.3, Lemma IV.4, and Corollary IV.5 we obtain the result of Theorem IV.1. Notice that the solution of Problem II.1 finds the partition of Σ into disjoint sets $\{\Sigma^0,\Sigma_0,\Sigma_1,\ldots,\Sigma_k\}$, where $\Sigma=\Sigma^0\cup\Sigma_0\cup\Sigma_1\cup\ldots\cup\Sigma_k$, and the optimal measure $\nu^*(\cdot)$ on these sets.

B. *Problem II.2: Finite Alphabet Case*

Consider Problem II.2, and follow the procedure utilized to derive the solution of Problem II.1 (e.g., Section IV-A). Let $\xi_i \triangleq \nu_i-\mu_i \equiv \xi_i^+ - \xi_i^-$, be the signed measure decomposition of ξ. We know that, $\sum_{i\in\Sigma}\xi_i=0$ and so, $\sum_{i\in\Sigma}\xi_i^+=\sum_{i\in\Sigma}\xi_i^-$. Also

$$\sum_{i\in\Sigma}|\nu_i-\mu_i| = \sum_{i\in\Sigma}|\xi_i| = \sum_{i\in\Sigma}\xi_i^+ + \sum_{i\in\Sigma}\xi_i^- = \alpha,\ \ \sum_{i\in\Sigma}\xi_i^+ = \sum_{i\in\Sigma}\xi_i^- = \frac{\alpha}{2}. \tag{49}$$

The average constraint can be written as follows

$$\sum_{i\in\Sigma}\ell_i\nu_i=\sum_{i\in\Sigma}\ell_i(\xi_i+\mu_i)=\sum_{i\in\Sigma}\ell_i\xi_i+\sum_{i\in\Sigma}\ell_i\mu_i=\sum_{i\in\Sigma}\ell_i\xi_i^+-\sum_{i\in\Sigma}\ell_i\xi_i^-+\sum_{i\in\Sigma}\ell_i\mu_i\leq D. \tag{50}$$

Define the maximum and minimum values of the sequence by $\ell_{\max}\triangleq\max_{i\in\Sigma}\ell_i$, $\ell_{\min}\triangleq\min_{i\in\Sigma}\ell_i$ and its corresponding support sets by $\Sigma^0\triangleq\{i\in\Sigma:\ell_i=\ell_{\max}\}$, $\Sigma_0\triangleq\{i\in\Sigma:\ell_i=\ell_{\min}\}$. For all remaining sequence, $\{\ell_i:i\in\Sigma\setminus\Sigma^0\cup\Sigma_0\}$, and for $1\leq r\leq|\Sigma\setminus\Sigma^0\cup\Sigma_0|$ define recursively

$$\Sigma^k\triangleq\left\{i\in\Sigma:\ell_i=\max\left\{\ell_\alpha:\alpha\in\Sigma\setminus\Sigma_0\cup\left(\bigcup_{j=1}^{k}\Sigma^{j-1}\right)\right\}\right\},\qquad k\in\{1,2,\ldots,r\}, \tag{51}$$

till all the elements of Σ are exhausted, and define the corresponding maximum value of ℓ on the sequence on these sets by

$$\ell\left(\Sigma^k\right)\triangleq\max_{i\in\Sigma\setminus\Sigma_0\cup\left(\bigcup_{j=1}^{k}\Sigma^{j-1}\right)}\ell_i,\qquad k\in\{1,2,\ldots,r\},$$

where r is the number of Σ^k sets which is at most $|\Sigma\setminus\Sigma^0\cup\Sigma_0|$. Clearly, $\ell\left(\Sigma^1\right)=\max_{i\in\Sigma\setminus\Sigma^0\cup\Sigma_0}\ell_i$ and so on. Note the analogy between (51) and (29) for Problem II.1. The main theorem which characterizes the extremum solution of Problem II.2 is given below.

Theorem IV.6. *The solution of the finite alphabet version of Problem II.2 is given by*

$$R^-(D)=\sum_{i\in\Sigma}|\nu_i^*-\mu_i|, \tag{52}$$

where the value of $R^-(D)$ is calculated as follows.

(1) *If*

$$\ell_{\min}\left(\sum_{j=0}^{k}\sum_{i\in\Sigma^j}\mu_i+\sum_{i\in\Sigma_0}\mu_i\right)+\sum_{j=k+1}^{r}\sum_{i\in\Sigma^j}\ell_i\mu_i\leq D\leq\ell_{\min}\left(\sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\mu_i+\sum_{i\in\Sigma_0}\mu_i\right)+\sum_{j=k}^{r}\sum_{i\in\Sigma^j}\ell_i\mu_i$$

then

$$R^-(D)=\frac{2\left(D-\ell_{\min}\sum_{i\in\Sigma_0}\mu_i-\ell\left(\Sigma^k\right)\sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\mu_i-\sum_{j=k}^{r}\sum_{i\in\Sigma^j}\ell_i\mu_i\right)}{\ell_{\min}-\ell\left(\Sigma^k\right)}. \tag{53}$$

(2) *If $D\geq(\ell_{\min}-\ell_{\max})\sum_{i\in\Sigma^0}\mu_i+\sum_{i\in\Sigma}\ell_i\mu_i$ then*

$$R^-(D)=\frac{2\left(D-\sum_{i\in\Sigma}\ell_i\mu_i\right)}{\ell_{\min}-\ell_{\max}}. \tag{54}$$

Moreover, the optimal probabilities are given by

$$\nu^*(\Sigma_0) \triangleq \sum_{i\in\Sigma_0} \nu_i^* = \sum_{i\in\Sigma_0} \mu_i + \alpha, \tag{55a}$$

$$\nu^*(\Sigma^0) \triangleq \sum_{i\in\Sigma^0} \nu_i^* = \left(\sum_{i\in\Sigma^0} \mu_i - \alpha\right)^+, \tag{55b}$$

$$\nu^*(\Sigma^k) \triangleq \sum_{i\in\Sigma^k} \nu_i^* = \left(\sum_{i\in\Sigma^k} \mu_i - \left(\alpha - \sum_{j=1}^{k} \sum_{i\in\Sigma^{j-1}} \mu_i\right)^+\right)^+, \tag{55c}$$

$$\alpha = \min\left(\frac{R^-(D)}{2}, 1 - \sum_{i\in\Sigma_0} \mu_i\right). \tag{55d}$$

where $k = 1,2,\ldots,r$ *and* r *is the number of* Σ^k *sets which is at most* $|\Sigma \setminus \Sigma^0 \cup \Sigma_0|$.

Proof: For the derivation of the Theorem see Appendix A. ■

C. Solutions of Related Extremum Problems

In Section II-A we discuss related extremum problems, whose solution can be obtained from those of Problem II.1 and Problem II.2. In this Section we give the solution of the finite alphabet version of the related extremum problems described by (8) and (10).

Consider the finite alphabet version of (8), that is

$$R^+(D) \stackrel{\triangle}{=} \sup_{\nu\in\mathcal{M}_1(\Sigma):\sum_{i\in\Sigma}\ell_i\nu_i\le D} ||\nu-\mu||_{TV}. \tag{56}$$

The solution of (56) is obtained from the solution of Problem II.1, by finding the inverse mapping or by following a similar procedure to the one utilized to derive Theorem IV.6.

Theorem IV.7. *The solution of the finite alphabet version of (56) is given by*

$$R^+(D) = \sum_{i\in\Sigma} |\nu_i^* - \mu_i|, \tag{57}$$

where the value of $R^+(D)$ *is calculated as follows.*

(1) *If*

$$\ell_{\max}\left(\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i + \sum_{i\in\Sigma^0}\mu_i\right) + \sum_{j=k}^{r}\sum_{i\in\Sigma_j}\ell_i\mu_i \le D \le \ell_{\max}\left(\sum_{j=0}^{k}\sum_{i\in\Sigma_j}\mu_i + \sum_{i\in\Sigma^0}\mu_i\right) + \sum_{j=k+1}^{r}\sum_{i\in\Sigma_j}\ell_i\mu_i$$

then

$$R^+(D) = \frac{2\left(D - \ell_{\max}\sum_{i\in\Sigma^0}\mu_i - \ell(\Sigma_k)\sum_{j=1}^{k}\sum_{i\in\Sigma_{j-1}}\mu_i - \sum_{j=k}^{r}\sum_{i\in\Sigma_j}\ell_i\mu_i\right)}{\ell_{\max} - \ell(\Sigma_k)}. \tag{58}$$

(2) *If* $D \leq (\ell_{\max} - \ell_{\min}) \sum_{i \in \Sigma_0} \mu_i + \sum_{i \in \Sigma} \ell_i \mu_i$ *then*

$$R^+(D) = \frac{2\left(D - \sum_{i \in \Sigma} \ell_i \mu_i\right)}{\ell_{\max} - \ell_{\min}}. \tag{59}$$

Moreover, the optimal probabilities are given by

$$\nu^*(\Sigma^0) \triangleq \sum_{i \in \Sigma^0} \nu_i^* = \sum_{i \in \Sigma^0} \mu_i + \alpha, \tag{60a}$$

$$\nu^*(\Sigma_0) \triangleq \sum_{i \in \Sigma_0} \nu_i^* = \left(\sum_{i \in \Sigma_0} \mu_i - \alpha\right)^+, \tag{60b}$$

$$\nu^*(\Sigma_k) \triangleq \sum_{i \in \Sigma_k} \nu_i^* = \left(\sum_{i \in \Sigma_k} \mu_i - \left(\alpha - \sum_{j=1}^{k} \sum_{i \in \Sigma_{j-1}} \mu_i\right)^+\right)^+, \tag{60c}$$

$$\alpha = \min\left(\frac{R^+(D)}{2}, 1 - \sum_{i \in \Sigma^0} \mu_i\right). \tag{60d}$$

where, $k = 1, 2, \ldots, r$ *and* r *is the number of* Σ_k *sets which is at most* $|\Sigma \setminus \Sigma^0 \cup \Sigma_0|$.

Consider the finite alphabet version of (10), that is

$$D^-(R) \triangleq \inf_{\nu \in \mathscr{M}_1(\Sigma): ||\nu - \mu||_{TV} \leq R} \sum_{i \in \Sigma} \ell_i \nu_i. \tag{61}$$

The solution of (61) is obtained from that of Problem II.1, but with a reverse computation on the partition of Σ and the mass of the extremum measure on the partition moving in the opposite direction. Below, we give the main theorem.

Theorem IV.8. *The solution of the finite alphabet version of (61) is given by*

$$D^-(R) = \ell_{\max} \nu^*(\Sigma^0) + \ell_{\min} \nu^*(\Sigma_0) + \sum_{k=1}^{r} \ell(\Sigma^k) \nu^*(\Sigma^k). \tag{62}$$

Moreover, the optimal probabilities are given by

$$\nu^*(\Sigma_0) \triangleq \sum_{i\in\Sigma_0} \nu_i^* = \sum_{i\in\Sigma_0} \mu_i + \alpha, \tag{63a}$$

$$\nu^*(\Sigma^0) \triangleq \sum_{i\in\Sigma^0} \nu_i^* = \left(\sum_{i\in\Sigma^0} \mu_i - \alpha\right)^+, \tag{63b}$$

$$\nu^*(\Sigma^k) \triangleq \sum_{i\in\Sigma^k} \nu_i^* = \left(\sum_{i\in\Sigma^k} \mu_i - \left(\alpha - \sum_{j=1}^{k} \sum_{i\in\Sigma^{j-1}} \mu_i\right)^+\right)^+, \tag{63c}$$

$$\alpha = \min\left(\frac{R}{2}, 1 - \sum_{i\in\Sigma_0} \mu_i\right), \tag{63d}$$

where, $k = 1,2,\ldots,r$ and r is the number of Σ^k sets which is at most $|\Sigma \setminus \Sigma^0 \cup \Sigma_0|$.

Remark IV.9. *The statements of Theorems IV.1, IV.6, IV.7, IV.8 are also valid for the countable alphabet case, because their derivations are not restricted to Σ being finite alphabet. It also holds for any $\ell \in BC^+(\Sigma)$ as seen in Section III. The extensions of Theorems IV.1-IV.8 to $\ell \in L^{\infty,+}(\Sigma, \mathscr{B}(\Sigma), \nu)$ can be shown as well; for example, $D^+(R)$ is given by*

$$D^+(R) = \ell_{\max} \nu^*(\Sigma^0) + \ell_{\min} \nu^*(\Sigma_0) + \sum_{k=1}^{r} \ell(\Sigma_k) \nu^*(\Sigma_k), \tag{64}$$

where the optimal probabilities are given by

$$\nu^*(\Sigma^0) = \mu(\Sigma^0) + \alpha, \tag{65a}$$

$$\nu^*(\Sigma_0) = (\mu(\Sigma_0) - \alpha)^+, \tag{65b}$$

$$\nu^*(\Sigma_k) = \left(\mu(\Sigma_k) - \left(\alpha - \sum_{j=1}^{k} \mu(\Sigma_{j-1})\right)^+\right)^+, \tag{65c}$$

$$\alpha = \min\left(\frac{R}{2}, 1 - \mu(\Sigma^0)\right), \tag{65d}$$

k is at most countable. We outline the main steps of the derivation. For any $n \in \mathbb{N}$, $\ell \in BC^+(\Sigma)$ define $\ell_n \triangleq \ell \wedge n$ (i.e., the minimum between ℓ and n), then $\ell_n \in BC^+(\Sigma)$, and for any $\nu \in \mathbb{B}_R(\mu)$ we have

$$\sup_{\nu\in\mathbb{B}_R(\mu)} \int_\Sigma \ell_n(x) d\nu(x) = \frac{R}{2}\left(\sup_{x\in\Sigma} \ell_n(x) - \inf_{x\in\Sigma} \ell_n(x)\right) + \int_\Sigma \ell_n(x) \nu(dx).$$

For any $\nu \in \mathbb{B}_R(\mu)$*, we obtain the inequality*

$$\begin{aligned}\int_\Sigma \ell(x) d\nu(x) &= \sup_{n\in\mathbb{N}} \int_\Sigma \ell_n(x)\nu(dx)\\ &\leq \sup_{n\in\mathbb{N}} \sup_{\nu\in\mathbb{B}_R(\mu)} \int_\Sigma \ell_n(x)\nu(dx)\\ &= \sup_{n\in\mathbb{N}} \left\{ \frac{R}{2}\left(\sup_{x\in\Sigma}\ell_n(x) - \inf_{x\in\Sigma}\ell_n(x)\right) + \int_\Sigma \ell_n(x) d\mu_n(x)\right\}\\ &\leq \sup_{n\in\mathbb{N}} \left\{ \frac{R}{2}\left(\sup_{x\in\Sigma}\ell_n(x) - \inf_{x\in\Sigma}\ell_n(x)\right)\right\} + \int_\Sigma \ell(x) d\mu(x).\end{aligned}$$

Hence,

$$\sup_{\nu\in\mathbb{B}_R(\mu)} \int_\Sigma \ell(x) d\nu(x) \leq \frac{R}{2}\sup_{n\in\mathbb{N}}\left\{\sup_{x\in\Sigma}\ell_n(x) - \inf_{x\in\Sigma}\ell_n(x)\right\} + \int_\Sigma \ell(x) d\mu(x).$$

Similarly, we can show that

$$\sup_{\nu\in\mathbb{B}_R(\mu)} \int_\Sigma \ell(x) d\nu(x) \geq \frac{R}{2}\sup_{n\in\mathbb{N}}\left\{\sup_{x\in\Sigma}\ell_n(x) - \inf_{x\in\Sigma}\ell_n(x)\right\} + \int_\Sigma \ell(x) d\mu(x).$$

Hence,

$$\sup_{\nu\in\mathbb{B}_R(\mu)} \int_\Sigma \ell(x) d\nu(x) = \frac{R}{2}\sup_{n\in\mathbb{N}}\left\{\sup_{x\in\Sigma}\ell_n(x) - \inf_{x\in\Sigma}\ell_n(x)\right\} + \int_\Sigma \ell(x) d\mu(x).$$

Utilizing the fact that $\sup_{n\in\mathbb{N}}\sup_{x\in\Sigma}\ell_n = \sup_n ||\ell_n||_{\infty,\nu}$ *(* $||\ell||_{\infty,\nu} = \inf_{\Delta\in N}\sup_{x\in\Delta^c}\ell(x)$, $N \triangleq \{A \in \mathscr{B}(\Sigma) : \nu(A) = 0\}$*, and similarly for the infimum) we obtain the results.*

V. Relation of Total Variational Distance to Other Metrics

In this section, we discuss relations of the total variational distance to other distance metrics. We also refer to some applications with distance metrics that can be substituted by the total variational distance metric.

L_1 *Distance Uncertainty.* Let $\sigma \in \mathscr{M}_1(\Sigma)$ be a fixed measure (as well as $\mu \in \mathscr{M}_1(\Sigma)$). Define the Radon-Nykodym derivatives $\psi \stackrel{\triangle}{=} \frac{d\mu}{d\sigma}$, $\varphi \stackrel{\triangle}{=} \frac{d\nu}{d\sigma}$ (densities with respect to a fixed $\sigma \in \mathscr{M}_1(\Sigma)$). Then,

$$||\nu - \mu||_{TV} = \int |\varphi(x) - \psi(x)|\sigma(dx)\;.$$

Consider a subset of $\mathbb{B}_R(\mu)$ defined by $\mathbb{B}_{R,\sigma}(\mu) \stackrel{\triangle}{=} \{\nu \in \mathbb{B}_R(\mu) : \nu << \sigma, \mu << \sigma\} \subseteq \mathbb{B}_R(\mu)$. Then,

$$\mathbb{B}_{R,\sigma}(\mu) = \left\{\varphi \in L_1(\sigma), \varphi \geq 0, \sigma - a.s. : \int_\Sigma |\varphi(x) - \psi(x)|\sigma(dx) \leq R\right\}\;.$$

Thus, under the absolute continuity of measures the total variational distance reduces to L_1 distance. Robustness via L_1 distance uncertainty on the space of spectral densities is investigated in the context of Wiener-Kolmogorov theory in an estimation and decision framework in [12], [13]. The extremum problem described under (a) can be applied to abstract formulations of minimax control and estimation, when the nominal system and uncertainty set are described by spectral measures with respect to variational distance.

Relative Entropy Uncertainty Model. [4] The relative entropy of $\nu \in \mathscr{M}_1(\Sigma)$ with respect to $\mu \in \mathscr{M}_1(\Sigma)$ is a mapping $H(\cdot|\cdot) : \mathscr{M}_1(\Sigma) \times \mathscr{M}_1(\Sigma) \longmapsto [0,\infty]$ defined by

$$H(\nu|\mu) \stackrel{\triangle}{=} \begin{cases} \int_\Sigma \log(\frac{d\nu}{d\mu})d\nu, & \text{if } \nu << \mu \\ +\infty, & \text{otherwise.} \end{cases}$$

It is well known that $H(\nu|\mu) \geq 0, \forall \nu, \mu \in \mathscr{M}_1(\Sigma)$, while $H(\nu|\mu) = 0 \Leftrightarrow \nu = \mu$. Total variational distance is bounded above by relative entropy via Pinsker's inequality giving

$$||\nu - \mu||_{TV} \leq \sqrt{2H(\nu|\mu)}, \quad \nu, \mu \in \mathscr{M}_1(\Sigma). \tag{66}$$

Given a known or nominal probability measure $\mu \in \mathscr{M}_1(\Sigma)$ the uncertainty set based on relative entropy is defined by $A_{\tilde{R}}(\mu) \stackrel{\triangle}{=} \{\nu \in \mathscr{M}_1(\Sigma) : H(\nu|\mu) \leq \tilde{R}\}$, where $\tilde{R} \in [0,\infty)$. Clearly, the uncertainty set determined by the total variation distance d_{TV}, is larger than that determined by the relative entropy. In other words, for every $r > 0$, in view of Pinsker's inequality (66):

$$\left\{\nu \in \mathscr{M}_1(\Sigma), \nu << \mu : H(\nu|\mu) \leq \frac{r^2}{2}\right\} \subseteq \mathbb{B}_R(\mu) \equiv \left\{\nu \in \mathscr{M}_1(\Sigma) : ||\nu - \mu||_{TV} \leq r\right\}.$$

Hence, even for those measures which satisfy $\nu << \mu$, the uncertainty set described by relative entropy is a subset of the much larger total variation distance uncertainty set. Moreover, by Pinsker's inequality, distance in total variation of probability measures is a lower bound on their relative entropy or Kullback-Leibler distance, and hence convergence in relative entropy of probability measures implies their convergence in total variation distance.

Over the last few years, relative entropy uncertainty model has received particular attention due to various properties (convexity, compact level sets), its simplicity and its connection to risk sensitive pay-off, minimax games, and large deviations [7]–[11]. Recently, an uncertainty model along the spirit of Radon-Nikodym derivative is employed in [27] for portfolio optimization under uncertainty. Unfortunately, relative entropy uncertainty modeling has two disadvantages. 1) It does not define a true metric on the space of measures; 2) relative entropy between two measures

is not defined if the measures are not absolutely continuous. The latter rules out the possibility of measures $\nu \in \mathscr{M}_1(\Sigma)$ and $\mu \in \mathscr{M}_1(\Sigma)$, $\tilde{\Sigma} \subset \Sigma$ to be defined on different spaces[4]. It is one of the main disadvantages in employing relative entropy in the context of uncertainty modelling for stochastic controlled diffusions (or SDE's) [28]. Specifically, by invoking a change of measure it can be shown that relative entropy modelling allows uncertainty in the drift coefficient of stochastic controlled diffusions, but not in the diffusion coefficient, because the latter kind of uncertainty leads to measures which are not absolutely continuous with respect to the nominal measure [7].

Kakutani-Hellinger Distance. [3] Another measure of distance of two probability measures which relates to their distance in variation is the Kakutani-Hellinger distance. Consider as before, $\nu \in \mathscr{M}_1(\Sigma)$, $\mu \in \mathscr{M}_1(\Sigma)$ and a fixed measure $\sigma \in \mathscr{M}_1(\Sigma)$ such that $\nu << \sigma$, $\mu << \sigma$ and define $\varphi \triangleq \frac{d\nu}{d\sigma}$, $\psi \triangleq \frac{d\mu}{d\sigma}$. The Kakutani-Hellinger distance is a mapping $d_{KH} : L_1(\sigma) \times L_1(\sigma) \mapsto [0,\infty)$ defined by

$$d_{KH}^2(\nu,\mu) \triangleq \frac{1}{2} \int \left(\sqrt{\varphi(x)} - \sqrt{\psi(x)} \right)^2 d\sigma(x). \tag{67}$$

Indeed, the function d_{KH} given by (67) is a metric on the set of probability measures. A related quantity is the Hellinger integral of measures $\nu \in \mathscr{M}_1(\Sigma)$ and $\mu \in \mathscr{M}_1(\Sigma)$ defined by

$$H(\nu,\mu) \triangleq \int \sqrt{\varphi(x)\psi(x)} d\sigma(x), \tag{68}$$

which is related to the Kakutani-Hellinger distance via $d_{KH}^2(\nu,\mu) = 1 - H(\nu,\mu)$. The relations between distance in variation and Kakutani-Hellinger distance (and Hellinger integral) are given by the following inequalities:

$$2\{1 - H(\nu,\mu)\} \leq ||\nu - \mu||_{TV} \leq \sqrt{8\{1 - H(\nu,\mu)\}}, \tag{69}$$

$$||\nu - \mu||_{TV} \leq 2\sqrt{1 - H^2(\nu,\mu)}, \tag{70}$$

$$2d_{KH}^2(\nu,\mu) \leq ||\nu - \mu||_{TV} \leq \sqrt{8} d_{KH}(\nu,\mu). \tag{71}$$

The above inequalities imply that these distances define the same topology on the space of probability measure on $(\Sigma, \mathscr{B}(\Sigma))$. Specifically, convergence in total variation of probability measures defined on a metric space $(\Sigma, \mathscr{B}(\Sigma), d)$, implies their weak convergence with respect

[4]This corresponds to the case in which the nominal system is a simplified version of the true system and is defined on a lower dimension space.

to the Kakutani-Hellinger distance metric, [3]. In [16], the Hellinger distance on the space of spectral densities is used to define a pay-off subject to constraints in the context of approximation theory.

Levy-Prohorov Distance. [4] Given a metric space $(\Sigma, \mathscr{B}(\Sigma), d)$, and a family of probability measures $\mathscr{M}_1(\Sigma)$ on $(\Sigma, \mathscr{M}_1(\Sigma))$ it is possible to "metrize" weak convergence of probability measure, denoted by $P_n \xrightarrow{w} P$, where $\{P_n : n \in \mathbb{N}\} \subset \mathscr{M}_1(\Sigma)$, $P \in \mathscr{M}_1(\Sigma)$ via the so called Levy-Prohorov metric denoted by $d_{LP}(\nu, \mu)$. Thus, this metric is also a candidate for a measure of proximity between two probability measures. The Levi-Prohorov metric is related to distance in variation via the upper bound [3],

$$d_{LP}(\nu, \mu) \leq \min\{||\nu - \mu||_{TV}, 1\}, \quad \forall \ \ \nu \in \mathscr{M}_1(\Sigma), \mu \in \mathscr{M}_1(\Sigma).$$

The function defined by $L(\nu, \mu) = \max\{d_{LP}(\nu, \mu), d_{LP}(\mu, \nu)\}$, is actually a distance metric (it satisfies the properties of distance).

In view of the relations between different metrics, such as relative entropy, Levy-Prohorov metric, Kakutani-Hellinger metric, it is clear that the Problem discussed under (1)-(4) give sub-optimal solution to the same problem with distance in variation replaced by these metrics.

VI. Examples

We will illustrate through simple examples how the optimal solution of the different extremum problems behaves. In particular, we present calculations through Example VI-A for $D^+(R)$ and $R^+(D)$, when the sequence $\ell = \{\ell_1 \ \ \ell_2 \ \ \ldots \ \ \ell_n\} \in \mathbb{R}^n_+$ consists of a number of ℓ_i's which are equal and calculations through Example VI-B for $R^-(D)$ and $D^-(R)$ when the ℓ_i's are not equal. We further present calculations through Example VI-C for $D^+(R)$, $R^+(D)$ and $D^-(R)$, $R^-(D)$ using a large number of ℓ_i's.

A. Example A

Let $\Sigma = \{i : i = 1, 2, \ldots, 8\}$ and for simplicity consider a descending sequence of lengths $\ell = \{\ell \in \mathbb{R}^8_+ : \ell_1 = \ell_2 > \ell_3 = \ell_4 > \ell_5 > \ell_6 = \ell_7 > \ell_8\}$ with corresponding nominal probability vector $\mu \in \mathbb{P}_1(\Sigma)$. Specifically, let $\ell = [1, 1, 0.8, 0.8, 0.6, 0.4, 0.4, 0.2]$, and $\mu = \left[\frac{23}{72}, \frac{13}{72}, \frac{10}{72}, \frac{9}{72}, \frac{8}{72}, \frac{4}{72}, \frac{3}{72}, \frac{2}{72}\right]$. Note that, the sets which correspond to the maximum, minimum and all the remaining lengths are equal to $\Sigma^0 = \{1, 2\}, \Sigma_0 = \{8\}, \Sigma_1 = \{7, 6\}, \Sigma_2 = \{5\}, \Sigma_3 = \{4, 3\}$. Figures 1(a)-(b) depicts

the maximum linear functional pay-off subject to total variational constraint, $D^+(R)$, and the optimal probabilities, both given by Theorem IV.1. Figures 1(c)-(d) depicts the maximum total variational pay-off subject to linear functional constraint, $R^+(D)$, and the optimal probabilities, both given by Theorem IV.7. Recall Lemma III.1 case 1 and Corollary III.3. Figure 1a shows that, $D^+(R)$ is a non-decreasing concave function of R and also that is constant in $[R_{\max}, 2]$, where $R_{\max} = 2\left(1-\mu(\Sigma^0)\right) = 1$.

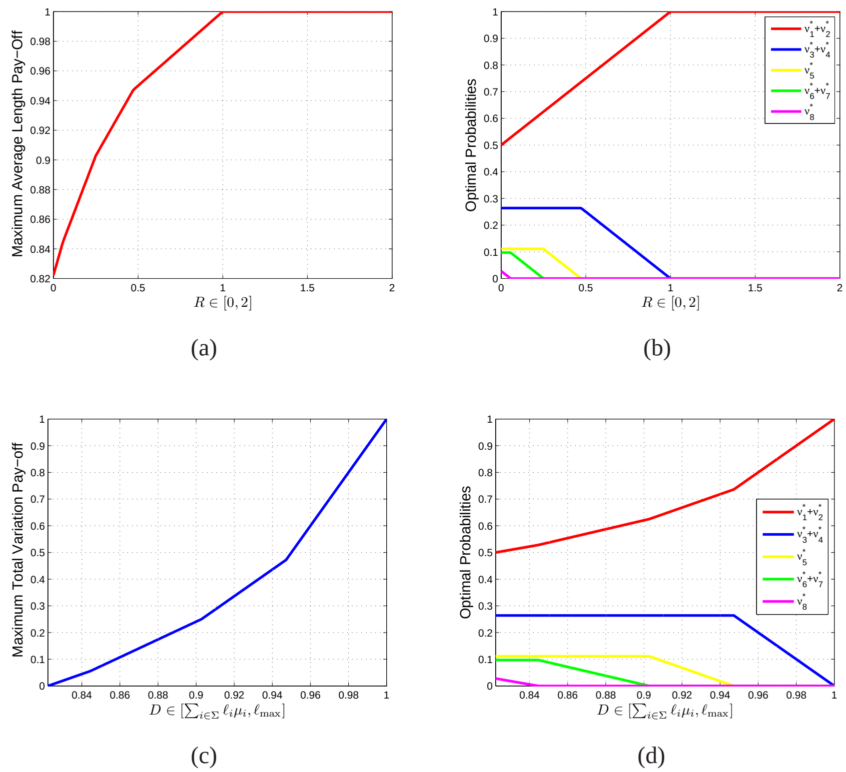

Fig. 1: Solution of Example A: (a) Optimum linear functional pay-off subject to total variational constraint, $D^+(R)$; (b) Optimal probabilities of $D^+(R)$; (c) Optimum total variational pay-off subject to linear functional constraint, $R^+(D)$; and, (d) Optimal probabilities of $R^+(D)$.

B. *Example B*

Let $\Sigma = \{i : i = 1,2,\ldots,8\}$ and for simplicity consider a descending sequence of lengths $\ell = \{\ell \in \mathbb{R}^8_+ : \ell_1 > \ell_2 > \ell_3 > \ell_4 > \ell_5 > \ell_6 > \ell_7 > \ell_8\}$ with corresponding nominal probability vector $\mu \in \mathbb{P}_1(\Sigma)$. Specifically, let $\ell = [1, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2]$ and $\mu = \left[\frac{23}{72}, \frac{13}{72}, \frac{10}{72}, \frac{9}{72}, \frac{8}{72}, \frac{4}{72}, \frac{3}{72}, \frac{2}{72}\right]$.

Note that, the sets which correspond to the maximum, minimum and all the remaining lengths are equal to $\Sigma^0 = \{1\}, \Sigma_0 = \{8\}, \Sigma^1 = \{2\}, \Sigma^2 = \{3\}, \Sigma^3 = \{4\}, \Sigma^4 = \{5\}, \Sigma^5 = \{6\}, \Sigma^6 = \{7\}$. Figures 2(a)-(b) depicts the minimum total variational pay-off subject to linear functional constraint, $R^-(D)$, and the optimal probabilities, both given by Theorem IV.6. Figures 2(c)-(d) depicts the minimum linear functional pay-off subject to total variational constraint, $D^-(R)$, and the optimal probabilities, both given by Theorem IV.8.

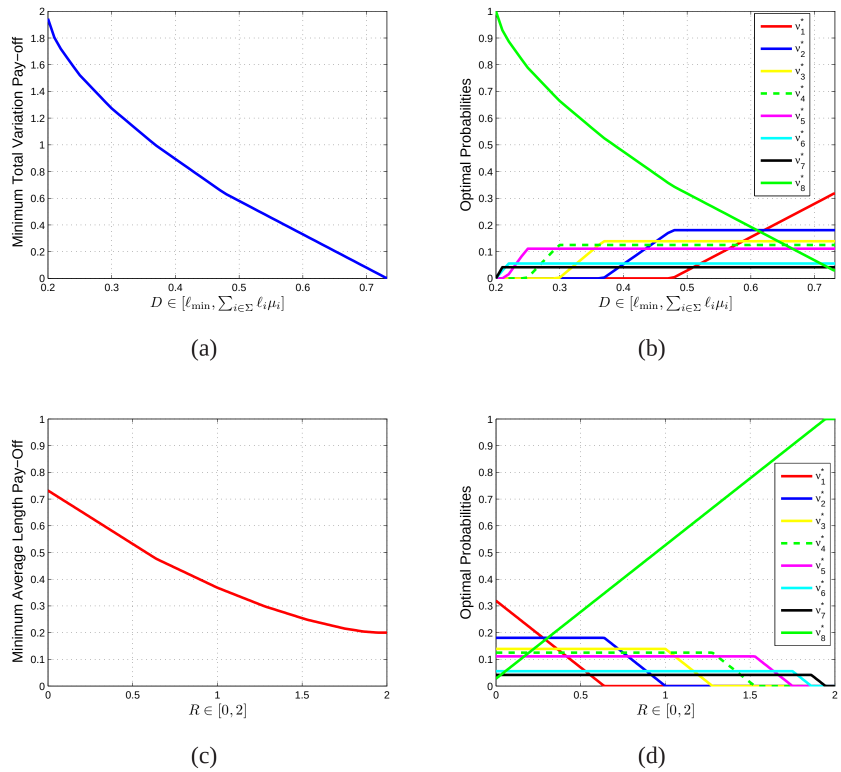

Fig. 2: Solution of Example B: (a) Optimum total variational pay-off subject to linear functional constraint, $R^-(D)$; (b) Optimal probabilities of $R^-(D)$; (c) Optimum linear functional pay-off subject to total variational constraint, $D^-(R)$; and, (d) Optimal probabilities of $D^-(R)$.

Recall Lemma III.1 case 2 and Corollary III.3. Figure 2a shows that, $R^-(D)$ is a non-increasing convex function of D, $D \in [\ell_{\min}, \sum_{i\in\Sigma} \ell_i \mu_i)$. Note that for $D < \ell_{\min} = 0.2$ no solution exists and $R^-(D)$ is zero in $[D_{\max}, \infty)$ where $D_{\max} = \sum_{i=1}^{8} \ell_i \mu_i = 0.73$.

C. Example C

Let $\Sigma = \{i : i = 1, 2, \ldots, 50\}$ and consider a descending sequence of lengths $\ell = \{\ell \in \mathbb{R}_+^{50}\}$ with corresponding nominal probability vector $\mu \in \mathbb{P}_1(\Sigma)$. For display purposes the support sets are denoted by Σ_x^y where $x, y = \{1, 2, \ldots, 16\}$, though of course the subscript symbol x corresponds to the support sets of Problem $D^+(R)$, $R^+(D)$ and the superscript symbol y corresponds to the support sets of Problem $D^-(R)$ and $R^-(D)$. Let

$$\ell = \Big[20\ \ 20\ \ 20\ \ 20\ \ 19\ \ 19\ \ 19\ \ 18\ \ 17\ \ 17\ \ 16\ \ 14\ \ 14\ \ 13\ \ 13\ \ 13\ \ 13\ \ 12\ \ 10\ \ 10\ \ 10\ \ 10$$
$$10\ \ 9\ \ 9\ \ 9\ \ 8\ \ 8\ \ 8\ \ 8\ \ 8\ \ 8\ \ 8\ \ 7\ \ 7\ \ 6\ \ 5\ \ 4\ \ 3\ \ 3\ \ 3\ \ 3\ \ 3\ \ 3\ \ 2\ \ 2\ \ 2\ \ 2\ \ 1\ \ 1\Big],$$

and

$$\mu = \Big[0.052\ \ 0.002\ \ 0.01\ \ 0.006\ \ 0.004\ \ 0.038\ \ 0.032\ \ 0.028\ \ 0.026\ \ 0.008\ \ 0.012\ \ 0.01\ \ 0.008$$
$$0.026\ \ 0.05\ \ 0.044\ \ 0.03\ \ 0.032\ \ 0.024\ \ 0.01\ \ 0.02\ \ 0.03\ \ 0.014\ \ 0.024\ \ 0.004\ \ 0.006\ \ 0.024$$
$$0.01\ \ 0.022\ \ 0.012\ \ 0.016\ \ 0.042\ \ 0.014\ \ 0.016\ \ 0.01\ \ 0.024\ \ 0.02\ \ 0.008\ \ 0.014\ \ 0.032\ \ 0.018$$
$$0.012\ \ \ 0.01\ \ \ 0.04\ \ \ 0.036\ \ \ 0.018\ \ \ 0.002\ \ \ 0.022\ \ \ 0.012\ \ \ 0.016\Big].$$

Note that, the sets which correspond to the maximum, minimum and all the remaining lengths are equal to

$$\Sigma^0 = \{1-4\}, \Sigma_0 = \{50, 49\}, \Sigma_1^{16} = \{48-45\}, \Sigma_2^{15} = \{44-39\}, \Sigma_3^{14} = \{38\}, \Sigma_4^{13} = \{37\},$$
$$\Sigma_5^{12} = \{36\}, \Sigma_6^{11} = \{35, 34\}, \Sigma_7^{10} = \{33-27\}, \Sigma_8^{9} = \{26-24\}, \Sigma_9^{8} = \{23-19\}, \Sigma_{10}^{7} = \{18\},$$
$$\Sigma_{11}^{6} = \{17-14\}, \Sigma_{12}^{5} = \{13, 12\}, \Sigma_{13}^{4} = \{11\}, \Sigma_{14}^{3} = \{10-9\}, \Sigma_{15}^{2} = \{8\}, \Sigma_{16}^{1} = \{7-5\}.$$

Figures 3(a)-(b) depicts the maximum linear functional pay-off subject to total variational constraint, $D^+(R)$, and the maximum total variational pay-off subject to linear functional constraint, $R^+(D)$, given by Theorem IV.1, IV.7, respectively. Figures 3(c)-(d) depicts the minimum linear functional pay-off subject to total variational constraint, $D^-(R)$, and the minimum total variational pay-off subject to linear functional constraint, $R^-(D)$, given by Theorem IV.8, IV.6 respectively.

VII. Conclusion

This paper is concerned with extremum problems involving total variational distance metric as a pay-off subject to linear functional constraints, and vice-versa; that is, with the roles of total

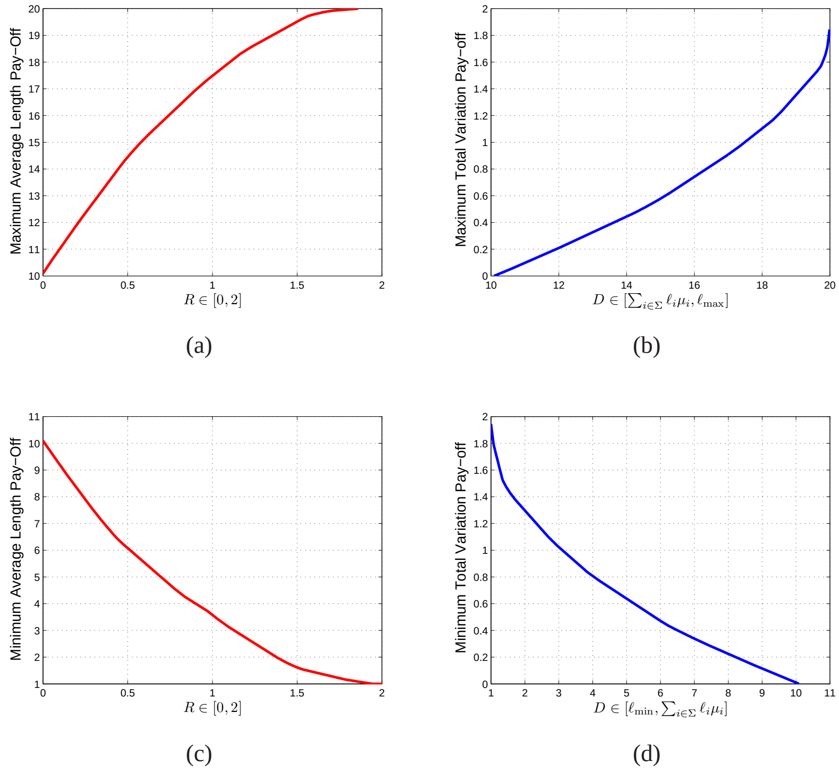

Fig. 3: Solution of Example C: (a) Optimum linear functional pay-off subject to total variational constraint, $D^+(R)$; (b) Optimum total variational pay-off subject to linear functional constraint, $R^+(D)$; (c) Optimum linear functional pay-off subject to total variational constraint, $D^-(R)$; and, (d) Optimum total variational pay-off subject to linear functional constraint, $R^-(D)$.

variational metric and linear functional interchanged. These problems are formulated using concepts from signed measures while the theory is developed on abstract spaces. Certain properties and applications of the extremum problems are discussed, while closed form expressions of the extremum measures are derived for finite alphabet spaces. Finally, it is shown through examples how the extremum solution of the various problems behaves. Extremum problems have a wide variety of applications, spanning from Markov decision problems to model reduction.

Appendix

Proof of Theorem IV.6

Lemma A.1. *The following bounds hold.*

1. Lower Bound.

$$\sum_{i\in\Sigma} \ell_i \xi_i^+ \geq \ell_{\min}\left(\frac{\alpha}{2}\right). \tag{72}$$

The bound holds with equality if

$$\sum_{i\in\Sigma_0} \mu_i + \frac{\alpha}{2} \leq 1, \quad \sum_{i\in\Sigma_0} \xi_i^+ = \frac{\alpha}{2}, \quad \xi_i^+ = 0 \;\; \textit{for} \;\; i \in \Sigma\setminus\Sigma_0,$$

and the optimal probability on Σ_0 is given by

$$\nu^*(\Sigma_0) \triangleq \sum_{i\in\Sigma_0} \nu_i^* = \min\left(1, \sum_{i\in\Sigma_0} \mu_i + \frac{\alpha}{2}\right).$$

2. Upper Bound.

$$\sum_{i\in\Sigma} \ell_i \xi_i^- \leq \ell_{\max}\left(\frac{\alpha}{2}\right). \tag{73}$$

The bound holds with equality if

$$\sum_{i\in\Sigma^0} \mu_i - \frac{\alpha}{2} \geq 0, \quad \sum_{i\in\Sigma^0} \xi_i^- = \frac{\alpha}{2}, \quad \xi_i^- = 0 \;\; \textit{for} \;\; i \in \Sigma\setminus\Sigma^0,$$

and the optimal probability on Σ^0 is given by

$$\nu^*(\Sigma^0) \triangleq \sum_{i\in\Sigma^0} \nu_i^* = \left(\sum_{i\in\Sigma^0} \mu_i - \frac{\alpha}{2}\right)^+.$$

Proof: Follows from Section III-A. ∎

Proposition A.2. *If $\sum_{i\in\Sigma_0}\mu_i + \frac{\alpha}{2} = 1$ and $\nu_i^* \geq \mu_i$ for all $i \in \Sigma_0$ then $R^-(D) = 2(1-\sum_{i\in\Sigma_0}\mu_i)$.*

Proof: The condition $\sum_{i\in\Sigma_0}\mu_i + \frac{\alpha}{2} = 1$ implies that $\sum_{i\in\Sigma_0}\nu_i^* = 1$ and therefore $\sum_{i\in\Sigma\setminus\Sigma_0}\nu_i^* = 0$, hence $\nu_i^* = 0$, for all $i \in \Sigma\setminus\Sigma_0$. Then the minimum pay-off (52) is given by

$$\begin{aligned}
R^-(D) &= \sum_{i\in\Sigma_0} |\nu_i^* - \mu_i| + \sum_{i\in\Sigma\setminus\Sigma_0} |\nu_i^* - \mu_i| = \sum_{i\in\Sigma_0} |\nu_i^* - \mu_i| + \sum_{i\in\Sigma\setminus\Sigma_0} |-\mu_i| \\
&\overset{(a)}{=} \sum_{i\in\Sigma_0} \nu_i^* - \sum_{i\in\Sigma_0} \mu_i + \sum_{i\in\Sigma\setminus\Sigma_0} \mu_i = \left(1 - \sum_{i\in\Sigma_0}\mu_i\right) + \left(1 - \sum_{i\in\Sigma_0}\mu_i\right) = 2\left(1 - \sum_{i\in\Sigma_0}\mu_i\right).
\end{aligned}$$

where (a) follows due to the fact that $\nu_i^* \geq \mu_i$ for all $i \in \Sigma_0$. ∎

Next, we show the derivation of (54).

Lemma A.3. *Under the conditions of Lemma A.1, then*

$$R^{-}(D) = \frac{2\left(D - \sum_{i\in\Sigma} \ell_i \mu_i\right)}{\ell_{\min} - \ell_{\max}}. \tag{74}$$

Proof: From (50) and Lemma A.1 we have

$$D \geq \sum_{i\in\Sigma} \ell_i \xi_i^{+} - \sum_{i\in\Sigma} \ell_i \xi_i^{-} + \sum_{i\in\Sigma} \ell_i \mu_i = \ell_{\min}\left(\frac{\alpha}{2}\right) - \ell_{\max}\left(\frac{\alpha}{2}\right) + \sum_{i\in\Sigma} \ell_i \mu_i.$$

Solving the above equation with respect to α we get that

$$\alpha \leq \frac{2\left(D - \sum_{i\in\Sigma} \ell_i \mu_i\right)}{\ell_{\min} - \ell_{\max}}.$$

If we select the solution on the boundary then, (74) is obtained. ∎

Corollary A.4. *For any $k \in \{1,2,\ldots,r\}$ if the following conditions hold*

$$\sum_{j=1}^{k} \sum_{i\in\Sigma^{j-1}} \mu_i - \frac{\alpha}{2} \leq 0 \;\; \text{and} \;\; \sum_{j=0}^{k} \sum_{i\in\Sigma^{j}} \mu_i - \frac{\alpha}{2} \geq 0, \tag{75a}$$

$$\sum_{i\in\Sigma^{j-1}} \xi_i^{-} = \sum_{i\in\Sigma^{j-1}} \mu_i, \;\; \text{for all} \;\; j = 1,2,\ldots,k, \tag{75b}$$

$$\sum_{i\in\Sigma^{k}} \xi_i^{-} = \left(\frac{\alpha}{2} - \sum_{j=1}^{k} \sum_{i\in\Sigma^{j-1}} \mu_i\right), \tag{75c}$$

$$\xi_i^{-} = 0 \;\; \text{for all} \;\; i \in \Sigma \setminus \Sigma^0 \cup \Sigma^1 \cup \ldots \cup \Sigma^k, \tag{75d}$$

$$\sum_{i\in\Sigma_0} \mu_i + \frac{\alpha}{2} < 1, \tag{75e}$$

$$\sum_{i\in\Sigma_0} \xi_i^{+} = \frac{\alpha}{2}, \quad \xi_i^{+} = 0 \;\; \text{for all} \;\; i \in \Sigma \setminus \Sigma_0, \tag{75f}$$

then

$$R^{-}(D) = \frac{2\left(D - \ell_{\min} \sum_{i\in\Sigma_0} \mu_i - \ell\left(\Sigma^k\right) \sum_{j=1}^{k} \sum_{i\in\Sigma^{j-1}} \mu_i - \sum_{j=k}^{r} \sum_{i\in\Sigma^{j}} \ell_i \mu_i\right)}{\ell_{\min} - \ell\left(\Sigma^k\right)}. \tag{76}$$

Moreover, the optimal probability on Σ^k is given by

$$\nu^*(\Sigma^k) \triangleq \sum_{i\in\Sigma^k} \nu_i^* = \left(\sum_{i\in\Sigma^k} \mu_i - \left(\frac{\alpha}{2} - \sum_{j=1}^{k} \sum_{i\in\Sigma^{j-1}} \mu_i\right)^{+}\right)^{+}. \tag{77}$$

Proof: Under the conditions stated, we have that

$$\sum_{i\in\Sigma}\ell_i\xi_i^- = \sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\ell_i\xi_i^- + \sum_{i\in\Sigma^k}\ell_i\xi_i^- + \sum_{i\in\Sigma\setminus\cup_{j=0}^{k}\Sigma^j}\ell_i\xi_i^- = \sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\ell_i\mu_i + \ell(\Sigma^k)\left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\mu_i\right).$$

Also,

$$\sum_{i\in\Sigma}\ell_i\xi_i^+ = \sum_{i\in\Sigma_0}\ell_i\xi_i^+ + \sum_{i\in\Sigma\setminus\Sigma_0}\ell_i\xi_i^+ = \ell_{\min}\left(\frac{\alpha}{2}\right).$$

From (50), we have that

$$\begin{aligned} D &\geq \sum_{i\in\Sigma}\ell_i\xi_i^+ - \sum_{i\in\Sigma}\ell_i\xi_i^- + \sum_{i\in\Sigma}\ell_i\mu_i \\ &= \ell_{\min}\left(\sum_{i\in\Sigma_0}\mu_i + \frac{\alpha}{2}\right) - \sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\ell_i\mu_i - \ell(\Sigma^k)\left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\mu_i\right) + \sum_{i\in\Sigma\setminus\Sigma_0}\ell_i\mu_i. \end{aligned}$$

Solving the above equation with respect to α we get that

$$\alpha \leq \frac{2\left(D - \ell_{\min}\sum_{i\in\Sigma_0}\mu_i - \ell(\Sigma^k)\sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\mu_i - \sum_{j=k}^{r}\sum_{i\in\Sigma^j}\ell_i\mu_i\right)}{\ell_{\min} - \ell(\Sigma^k)}.$$

If we select the solution at the boundary then, (76) is obtained. From (75c) we have that

$$\sum_{i\in\Sigma^k}\xi_i^- = \left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\mu_i\right), \text{ and hence, } \sum_{i\in\Sigma^k}\nu_i = \sum_{i\in\Sigma^k}\mu_i - \left(\frac{\alpha}{2} - \sum_{j=1}^{k}\sum_{i\in\Sigma^{j-1}}\mu_i\right).$$

The optimal $\sum_{i\in\Sigma^k}\nu_i^*$ must satisfy (75a). Hence, (77) is obtained. ■

Putting together Lemma A.1, Proposition A.2, Lemma A.3, and Corollary A.4 we obtain the result of Theorem IV.6.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, Inc., 1991.

[2] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. London: Springer-Verlag, 1993.

[3] A. L. Gibbs and F. E. SU, "On choosing and bounding probability metrics," *Internat. Statist. Rev*, vol. 70, no. 3, pp. 419–435, Dec. 2002.

[4] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. New York: John Wiley & Sons, Inc., 1997.

[5] C. D. Charalambous, I. Tzortzis, and F. Rezaei, "Stochastic optimal control of discrete-time systems subject to conditional distribution uncertainty," in *50th IEEE Conference on Decision and Control and European Control Conference*, Orlando, Florida, Dec. 12–15, 2011, pp. 6407–6412.

[6] N. Dunford and J. Schwartz, *Linear Operators: Part 1: General Theory*. New York: Interscience Publishers, Inc., 1957.

[7] P. D. Pra, L. Meneghini, and W. J. Runggaldier, "Connections between stochastic control and dynamic games," *Math. Control Signals Systems*, vol. 9, no. 4, pp. 303–326, 1996.

[8] V. A. Ugrinovskii and I. R. Petersen, “Finite horizon minimax optimal control of stochastic partially observed time varying uncertain systems,” *Math. Control Signals Systems*, vol. 12, no. 1, pp. 1–23, 1999.

[9] I. R. Petersen, M. R. James, and P. Dupuis, “Minimax optimal control of stochastic uncertain systems with relative entropy constraints,” *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 398–412, Mar. 2000.

[10] N. U. Ahmed and C. D. Charalambous, “Minimax games for stochastic systems subject to relative entropy uncertainty: Applications to sde’s on hilbert spaces,” *J. Math. Control Signals Systems*, vol. 19, no. 1, pp. 65–91, Feb. 2007.

[11] C. D. Charalambous and F. Rezaei, “Stochastic uncertain systems subject to relative entropy constraints: Induced norms and monotonicity properties of minimax games,” *IEEE Trans. Autom. Control*, vol. 52, no. 4, pp. 647–663, Apr. 2007.

[12] H. V. Poor, “On robust wiener filtering,” *IEEE Trans. Autom. Control*, vol. 25, no. 3, pp. 531–536, Jun. 1980.

[13] K. S. Vastola and H. V. Poor, “On robust wiener-kolmogorov theory,” *IEEE Trans. Inform. Theory*, vol. 30, no. 2, pp. 315–327, Mar. 1984.

[14] C. D. Charalambous, I. Tzortzis, and T. Charalambous, “Dynamic programming with total variational distance uncertainty,” in *51st IEEE Conference on Decision and Control*, Maui, Hawaii, Dec. 10–13, 2012.

[15] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems.* New York: Springer-Verlag, 1985.

[16] A. Ferrante, M. Pavon, and F. Ramponi, “Hellinger vs. kullback-leibler multivariable spectrum approximation,” *IEEE Trans. Autom. Control*, vol. 53, no. 5, pp. 954–967, May 2008.

[17] T. T. Georgiou, “Relative entropy and the multivariable multidimensional moment problem,” *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1052–1066, Mar. 2006.

[18] A. Ferrante, M. Pavon, and F. Ramponi, “Constrained approximation in the hellinger distance,” in *Proceedings of European Control Conference*, Kos, Greece, Jul. 2–5, 2007, pp. 322–327.

[19] T. T. Georgiou and A. Lindquist, “Kullback-leibler approximation of spectral density functions,” *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 2910–2917, Nov. 2003.

[20] M. Pavon and A. Ferrante, “On the georgiou-lindquist approach to constrained kullback-leibler approximation of spectral densities,” *IEEE Trans. Autom. Control*, vol. 51, no. 4, pp. 639–644, Apr. 2006.

[21] F. Rezaei, C. D. Charalambous, and N. U. Ahmed, “Optimal control of uncertain stochastic systems subject to total variation distance uncertainty,” *SIAM Journal on Control and Optimization*, vol. 50, no. 5, pp. 2683–2725, Sep. 2012.

[22] E. T. Jaynes, “Information theory and statistical mechanics,” *Physics Review*, vol. 106, pp. 620–630, 1957.

[23] ——, “Information theory and statistical mechanics ii,” *Physics Review*, vol. 108, pp. 171–190, 1957.

[24] J. S. Baras and M. Rabi, “Maximum entropy models, dynamic games, and robust output feedback control for automata,” in *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, Seville, Spain, Dec. 12–15, 2005.

[25] F. Rezaei, C. D. Charalambous, and N. U. Ahmed, *Optimization of Stochastic Uncertain Systems: Entropy Rate Functionals, Minimax Games and Robustness. A Festschrift in Honor of Robert J Elliott*, ser. Advances in Statistics, Probability and Actuarial Science. World Scientific Publishing Company Incorporated, 2012.

[26] P. R. Halmos, *Measure Theory*. Springer-Verlag New York Inc., 1974.

[27] B. Oksendal and A. Sulem, “Portfolio optimization under model uncertainty and bsde games,” Institut National de Recherche en Informatique et en Automatique, Rapport de Recherche 7554, 2011.

[28] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes.* San Francisco: Holden-Day, 1964.